



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁷ : C12N 15/31, C07K 14/315, 16/12, A61K 31/70, 39/09, G01N 33/53, 33/68, C12Q 1/68</p>	A2	<p>(11) International Publication Number: WO 00/06738</p> <p>(43) International Publication Date: 10 February 2000 (10.02.00)</p>
<p>(21) International Application Number: PCT/GB99/02452</p> <p>(22) International Filing Date: 27 July 1999 (27.07.99)</p> <p>(30) Priority Data: 9816336.3 27 July 1998 (27.07.98) GB 60/125,329 19 March 1999 (19.03.99) US</p> <p>(71) Applicant (for all designated States except US): MICROBIAL TECHNIQS LIMITED [GB/GB]; 20 Trumpington Street, Cambridge CB2 1QA (GB).</p> <p>(72) Inventors; and</p> <p>(75) Inventors/Applicants (for US only): LE PAGE, Richard, William, Falla [GB/GB]; University of Cambridge, Dept. of Pathology, Tennis Court Road, Cambridge CB2 1QP (GB). WELLS, Jeremy, Mark [GB/GB]; Actinova Ltd., 12 Pembroke Avenue, Denny End Industrial Centre, Waterbeech, Cambridge CB5 9PB (GB). HANNIFFY, Sean, Bosco [IE/GB]; University of Cambridge, Dept. of Pathology, Tennis Court Road, Cambridge CB2 1QP (GB). HANSBRO, Philip, Michael [GB/GB]; University of Cambridge, Dept. of Pathology, Tennis Court Road, Cambridge CB2 1QP (GB).</p>	<p>(74) Agents: CHAPMAN, Paul, William et al.; Kilburn & Strobe, 20 Red Lion Street, London WC1R 4PJ (GB).</p> <p>(81) Designated States: CN, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).</p> <p>Published Without international search report and to be republished upon receipt of that report.</p>	
<p>(54) Title: NUCLEIC ACIDS AND PROTEINS FROM STREPTOCOCCUS PNEUMONIAE</p> <p>(57) Abstract</p> <p>Novel proteins from <i>Streptococcus pneumoniae</i> are described, together with nucleic acid sequences encoding them. Their use in vaccines and in screening methods is also described.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

NUCLEIC ACIDS AND PROTEINS FROM STREPTOCOCCUS PNEUMONIAE

The present invention relates to proteins derived from *Streptococcus pneumoniae*, nucleic acid molecules encoding such proteins, the use of the nucleic acid and/or proteins as antigens/immunogens and in detection/diagnosis, as well as methods for screening the proteins/nucleic acid sequences as potential anti-microbial targets.

Streptococcus pneumoniae, commonly referred to as the pneumococcus, is an important pathogenic organism. The continuing significance of *Streptococcus pneumoniae* infections in relation to human disease in developing and developed countries has been authoritatively reviewed (Fiber, G.R., *Science*, **265**: 1385-1387 (1994)). That indicates that on a global scale this organism is believed to be the most common bacterial cause of acute respiratory infections, and is estimated to result in 1 million childhood deaths each year, mostly in developing countries (Stansfield, S.K., *Pediatr. Infect. Dis.*, **6**: 622 (1987)). In the USA it has been suggested (Breiman *et al.*, *Arch. Intern. Med.*, **150**: 1401 (1990)) that the pneumococcus is still the most common cause of bacterial pneumonia, and that disease rates are particularly high in young children, in the elderly, and in patients with predisposing conditions such as asplenia, heart, lung and kidney disease, diabetes, alcoholism, or with immunosuppressive disorders, especially AIDS. These groups are at higher risk of pneumococcal septicaemia and hence meningitis and therefore have a greater risk of dying from pneumococcal infection. The pneumococcus is also the leading cause of otitis media and sinusitis, which remain prevalent infections in children in developed countries, and which incur substantial costs.

25

The need for effective preventative strategies against pneumococcal infection is highlighted by the recent emergence of penicillin-resistant pneumococci. It has been reported that 6.6% of pneumococcal isolates in 13 US hospitals in 12 states were found

to be resistant to penicillin and some isolates were also resistant to other antibiotics including third generation cyclosporins (Schappert, S.M., *Vital and Health Statistics of the Centres for Disease Control/National Centre for Health Statistics*, 214:1 (1992)). The rates of penicillin resistance can be higher (up to 20%) in some hospitals (Breiman *et al*, J. Am. Med. Assoc., 271: 1831 (1994)). Since the development of penicillin resistance among pneumococci is both recent and sudden, coming after decades during which penicillin remained an effective treatment, these findings are regarded as alarming.

For the reasons given above, there are therefore compelling grounds for considering improvements in the means of preventing, controlling, diagnosing or treating pneumococcal diseases.

Various approaches have been taken in order to provide vaccines for the prevention of pneumococcal infections. Difficulties arise for instance in view of the variety of serotypes (at least 90) based on the structure of the polysaccharide capsule surrounding the organism. Vaccines against individual serotypes are not effective against other serotypes and this means that vaccines must include polysaccharide antigens from a whole range of serotypes in order to be effective in a majority of cases. An additional problem arises because it has been found that the capsular polysaccharides (each of which determines the serotype and is the major protective antigen) when purified and used as a vaccine do not reliably induce protective antibody responses in children under two years of age, the age group which suffers the highest incidence of invasive pneumococcal infection and meningitis.

A modification of the approach using capsule antigens relies on conjugating the polysaccharide to a protein in order to derive an enhanced immune response, particularly by giving the response T-cell dependent character. This approach has

been used in the development of a vaccine against *Haemophilus influenzae*, for instance. There are, however, issues of cost concerning both the multi-polysaccharide vaccines and those based on conjugates.

5 A third approach is to look for other antigenic components which offer the potential to be vaccine candidates. This is the basis of the present invention. Using a specially developed bacterial expression system, we have been able to identify a group of protein antigens from pneumococcus which are associated with the bacterial envelope or which are secreted.

10

Thus, in a first aspect the present invention provides a *Streptococcus pneumoniae* protein or polypeptide having a sequence selected from those shown in table 1.

15 In a second aspect, the present invention provides a *Streptococcus pneumoniae* protein or polypeptide having a sequence selected from those shown in table 2.

A protein or polypeptide of the present invention may be provided in substantially pure form. For example, it may be provided in a form which is substantially free of other proteins.

20

As discussed herein, the proteins and polypeptides of the invention are useful as antigenic material. Such material can be "antigenic" and/or "immunogenic". Generally, "antigenic" is taken to mean that the protein or polypeptide is capable of being used to raise antibodies or indeed is capable of inducing an antibody response in a subject. "Immunogenic" is taken to mean that the protein or polypeptide is capable of eliciting a protective immune response in a subject. Thus, in the latter case, the protein or polypeptide may be capable of not only generating an antibody response but, in addition, a non-antibody based immune response.

25

The skilled person will appreciate that homologues or derivatives of the proteins or polypeptides of the invention will also find use in the context of the present invention, ie as antigenic/immunogenic material. Thus, for instance proteins or polypeptides which include one or more additions, deletions, substitutions or the like are encompassed by the present invention. In addition, it may be possible to replace one amino acid with another of similar "type". For instance replacing one hydrophobic amino acid with another.

One can use a program such as the CLUSTAL program to compare amino acid sequences. This program compares amino acid sequences and finds the optimal alignment by inserting spaces in either sequence as appropriate. It is possible to calculate amino acid identity or similarity (identity plus conservation of amino acid type) for an optimal alignment. A program like BLASTx will align the longest stretch of similar sequences and assign a value to the fit. It is thus possible to obtain a comparison where several regions of similarity are found, each having a different score. Both types of identity analysis are contemplated in the present invention.

In the case of homologues and derivatives, the degree of identity with a protein or polypeptide as described herein is less important than that the homologue or derivative should retain the antigenicity or immunogenicity of the original protein or polypeptide. However, suitably, homologues or derivatives having at least 60% similarity (as discussed above) with the proteins or polypeptides described herein are provided. Preferably, homologues or derivatives having at least 70% similarity, more preferably at least 80% similarity are provided. Most preferably, homologues or derivatives having at least 90% or even 95% similarity are provided.

In an alternative approach, the homologues or derivatives could be fusion proteins, incorporating moieties which render purification easier, for example by effectively

tagging the desired protein or polypeptide. It may be necessary to remove the "tag" or it may be the case that the fusion protein itself retains sufficient antigenicity to be useful.

- 5 In an additional aspect of the invention there are provided antigenic/immunogenic fragments of the proteins or polypeptides of the invention, or of homologues or derivatives thereof.

10 For fragments of the proteins or polypeptides described herein, or of homologues or derivatives thereof, the situation is slightly different. It is well known that is possible to screen an antigenic protein or polypeptide to identify epitopic regions, ie those regions which are responsible for the protein or polypeptide's antigenicity or immunogenicity. Methods for carrying out such screening are well known in the art. Thus, the fragments of the present invention should include one or more such epitopic regions or be
15 sufficiently similar to such regions to retain their antigenic/immunogenic properties. Thus, for fragments according to the present invention the degree of identity is perhaps irrelevant, since they may be 100% identical to a particular part of a protein or polypeptide, homologue or derivative as described herein. The key issue, once again, is that the fragment retains the antigenic/immunogenic properties.

20

Thus, what is important for homologues, derivatives and fragments is that they possess at least a degree of the antigenicity/immunogenicity of the protein or polypeptide from which they are derived.

- 25 Gene cloning techniques may be used to provide a protein of the invention in substantially pure form. These techniques are disclosed, for example, in J. Sambrook *et al Molecular Cloning* 2nd Edition, Cold Spring Harbor Laboratory Press (1989). Thus, in a third aspect, the present invention provides a nucleic acid molecule

comprising or consisting of a sequence which is:

- (i) any of the DNA sequences set out in Table 1 or their RNA equivalents;
- 5 (ii) a sequence which is complementary to any of the sequences of (i);
- (iii) a sequence which codes for the same protein or polypeptide, as those sequences of (i) or (ii);
- 10 (iv) a sequence which has substantial identity with any of those of (i), (ii) and (iii);
- (v) a sequence which codes for a homologue, derivative or fragment of a protein as defined in Table 1.

15

In a fourth aspect the present invention provides a nucleic acid molecule comprising or consisting of a sequence which is:

- (i) any of the DNA sequences set out in Table 2 or their RNA equivalents;
- 20 (ii) a sequence which is complementary to any of the sequences of (i);
- (iii) a sequence which codes for the same protein or polypeptide, as those sequences of (i) or (ii);
- 25 (iv) a sequence which has substantial identity with any of those of (i), (ii) and (iii); or

- (v) a sequence which codes for a homologue, derivative or fragment of a protein as defined in Table 2.

5 The nucleic acid molecules of the invention may include a plurality of such sequences, and/or fragments. The skilled person will appreciate that the present invention can include novel variants of those particular novel nucleic acid molecules which are exemplified herein. Such variants are encompassed by the present invention. These may occur in nature, for example because of strain variation. For example, additions, substitutions and/or deletions are included. In addition, and particularly when utilising
10 microbial expression systems, one may wish to engineer the nucleic acid sequence by making use of known preferred codon usage in the particular organism being used for expression. Thus, synthetic or non-naturally occurring variants are also included within the scope of the invention.

15 The term "RNA equivalent" when used above indicates that a given RNA molecule has a sequence which is complementary to that of a given DNA molecule (allowing for the fact that in RNA "U" replaces "T" in the genetic code).

When comparing nucleic acid sequences for the purposes of determining the degree of
20 homology or identity one can use programs such as BESTFIT and GAP (both from the Wisconsin Genetics Computer Group (GCG) software package) BESTFIT, for example, compares two sequences and produces an optimal alignment of the most similar segments. GAP enables sequences to be aligned along their whole length and finds the optimal alignment by inserting spaces in either sequence as appropriate.
25 Suitably, in the context of the present invention when discussing identity of nucleic acid sequences, the comparison is made by alignment of the sequences along their whole length.

Preferably, sequences which have substantial identity have at least 50% sequence identity, desirably at least 75% sequence identity and more desirably at least 90 or at least 95% sequence identity with said sequences. In some cases the sequence identity may be 99% or above.

5

Desirably, the term "substantial identity" indicates that said sequence has a greater degree of identity with any of the sequences described herein than with prior art nucleic acid sequences.

10

It should however be noted that where a nucleic acid sequence of the present invention codes for at least part of a novel gene product the present invention includes within its scope all possible sequence coding for the gene product or for a novel part thereof.

15

The nucleic acid molecule may be in isolated or recombinant form. It may be incorporated into a vector and the vector may be incorporated into a host. Such vectors and suitable hosts form yet further aspects of the present invention.

20

Therefore, for example, by using probes based upon the nucleic acid sequences provided herein, genes in *Streptococcus pneumoniae* can be identified. They can then be excised using restriction enzymes and cloned into a vector. The vector can be introduced into a suitable host for expression.

25

Nucleic acid molecules of the present invention may be obtained from *S.pneumoniae* by the use of appropriate probes complementary to part of the sequences of the nucleic acid molecules. Restriction enzymes or sonication techniques can be used to obtain appropriately sized fragments for probing.

Alternatively PCR techniques may be used to amplify a desired nucleic acid sequence. Thus the sequence data provided herein can be used to design two primers for use in PCR so that a desired sequence, including whole genes or fragments thereof, can be targeted and then amplified to a high degree.

5

Typically primers will be at least 15-25 nucleotides long.

As a further alternative chemical synthesis may be used. This may be automated.

Relatively short sequences may be chemically synthesised and ligated together to provide a longer sequence.

10

There is another group of proteins from *S.pneumoniae* which have been identified using the bacterial expression system described herein. These are known proteins from *S.pneumoniae*, which have not previously been identified as antigenic proteins.

15

The amino acid sequences of this group of proteins, together with DNA sequences coding for them are shown in Table 3. These proteins, or homologues, derivatives and/or fragments thereof also find use as antigens/immunogens. Thus, in another aspect the present invention provides the use of a protein or polypeptide having a sequence selected from those shown in Tables 1-3, or homologues, derivatives and/or fragments thereof, as an immunogen/antigen.

20

In yet a further aspect the present invention provides an immunogenic/antigenic composition comprising one or more proteins or polypeptides selected from those whose sequences are shown in Tables 1-3, or homologues or derivatives thereof, and/or fragments of any of these. In preferred embodiments, the immunogenic/antigenic composition is a vaccine or is for use in a diagnostic assay.

25

In the case of vaccines suitable additional excipients, diluents, adjuvants or the like

may be included. Numerous examples of these are well known in the art.

It is also possible to utilise the nucleic acid sequences shown in Tables 1-3 in the preparation of so-called DNA vaccines. Thus, the invention also provides a vaccine
5 composition comprising one or more nucleic acid sequences as defined herein. DNA vaccines are described in the art (see for instance, Donnelly *et al*, *Ann. Rev. Immunol.*, 15:617-648 (1997)) and the skilled person can use such art described techniques to produce and use DNA vaccines according to the present invention.

10 As already discussed herein the proteins or polypeptides described herein, their homologues or derivatives, and/or fragments of any of these, can be used in methods of detecting/diagnosing *S.pneumoniae*. Such methods can be based on the detection of antibodies against such proteins which may be present in a subject. Therefore the present invention provides a method for the detection/diagnosis of *S.pneumoniae*
15 which comprises the step of bringing into contact a sample to be tested with at least one protein, or homologue, derivative or fragment thereof, as described herein. Suitably, the sample is a biological sample, such as a tissue sample or a sample of blood or saliva obtained from a subject to be tested.

20 In an alternative approach, the proteins described herein, or homologues, derivatives and/or fragments thereof, can be used to raise antibodies, which in turn can be used to detect the antigens, and hence *S.pneumoniae*. Such antibodies form another aspect of the invention. Antibodies within the scope of the present invention may be monoclonal or polyclonal.

25 Polyclonal antibodies can be raised by stimulating their production in a suitable animal host (e.g. a mouse, rat, guinea pig, rabbit, sheep, goat or monkey) when a protein as described herein, or a homologue, derivative or fragment thereof, is injected into the

animal. If desired, an adjuvant may be administered together with the protein. Well-known adjuvants include Freund's adjuvant (complete and incomplete) and aluminium hydroxide. The antibodies can then be purified by virtue of their binding to a protein as described herein.

5

Monoclonal antibodies can be produced from hybridomas. These can be formed by fusing myeloma cells and spleen cells which produce the desired antibody in order to form an immortal cell line. Thus the well-known Kohler & Milstein technique (*Nature* 256 (1975)) or subsequent variations upon this technique can be used.

10

Techniques for producing monoclonal and polyclonal antibodies that bind to a particular polypeptide/protein are now well developed in the art. They are discussed in standard immunology textbooks, for example in Roitt *et al*, *Immunology* second edition (1989), Churchill Livingstone, London.

15

In addition to whole antibodies, the present invention includes derivatives thereof which are capable of binding to proteins etc as described herein. Thus the present invention includes antibody fragments and synthetic constructs. Examples of antibody fragments and synthetic constructs are given by Dougall *et al* in *Tibtech* 12 372-379 (September 1994).

20

25

Antibody fragments include, for example, Fab, F(ab')₂ and Fv fragments. Fab fragments (These are discussed in Roitt *et al* [*supra*]). Fv fragments can be modified to produce a synthetic construct known as a single chain Fv (scFv) molecule. This includes a peptide linker covalently joining V_H and V_L regions, which contributes to the stability of the molecule. Other synthetic constructs that can be used include CDR peptides. These are synthetic peptides comprising antigen-binding determinants. Peptide mimetics may also be used. These molecules are usually conformationally

restricted organic rings that mimic the structure of a CDR loop and that include antigen-interactive side chains.

5 Synthetic constructs include chimaeric molecules. Thus, for example, humanised (or primatised) antibodies or derivatives thereof are within the scope of the present invention. An example of a humanised antibody is an antibody having human framework regions, but rodent hypervariable regions. Ways of producing chimaeric antibodies are discussed for example by Morrison *et al* in PNAS, 81, 6851-6855 (1984) and by Takeda *et al* in Nature. 314, 452-454 (1985).

10

Synthetic constructs also include molecules comprising an additional moiety that provides the molecule with some desirable property in addition to antigen binding. For example the moiety may be a label (e.g. a fluorescent or radioactive label). Alternatively, it may be a pharmaceutically active agent.

15

Antibodies, or derivatives thereof, find use in detection/diagnosis of *S.pneumoniae*. Thus, in another aspect the present invention provides a method for the detection/diagnosis of *S.pneumoniae* which comprises the step of bringing into contact a sample to be tested and antibodies capable of binding to one or more proteins described herein, or to homologues, derivatives and/or fragments thereof.

20

In addition, so-called "Affibodies" may be utilised. These are binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain (Nord *et al* ,) Thus, Small protein domains, capable of specific binding to different target proteins can be selected using combinatorial approaches.

25

It will also be clear that the nucleic acid sequences described herein may be used to detect/diagnose *S.pneumoniae*. Thus, in yet a further aspect, the present invention

provides a method for the detection/diagnosis of *S.pneumoniae* which comprises the step of bringing into contact a sample to be tested with at least one nucleic acid sequence as described herein. Suitably, the sample is a biological sample, such as a tissue sample or a sample of blood or saliva obtained from a subject to be tested.

- 5 Such samples may be pre-treated before being used in the methods of the invention. Thus, for example, a sample may be treated to extract DNA. Then, DNA probes based on the nucleic acid sequences described herein (ie usually fragments of such sequences) may be used to detect nucleic acid from *S.pneumoniae*.

- 10 In additional aspects, the present invention provides:

- (a) a method of vaccinating a subject against *S.pneumoniae* which comprises the step of administering to a subject a protein or polypeptide of the invention, or a derivative, homologue or fragment thereof, or an immunogenic composition of the
15 invention;
- (b) a method of vaccinating a subject against *S.pneumoniae* which comprises the step of administering to a subject a nucleic acid molecule as defined herein;
- 20 (c) a method for the prophylaxis or treatment of *S.pneumoniae* infection which comprises the step of administering to a subject a protein or polypeptide of the invention, or a derivative, homologue or fragment thereof, or an immunogenic composition of the invention;
- 25 (d) a method for the prophylaxis or treatment of *S.pneumoniae* infection which comprises the step of administering to a subject a nucleic acid molecule as defined herein;

- (e) a kit for use in detecting/diagnosing *S.pneumoniae* infection comprising one or more proteins or polypeptides of the invention, or homologues, derivatives or fragments thereof, or an antigenic composition of the invention; and
- 5 (f) a kit for use in detecting/diagnosing *S.pneumoniae* infection comprising one or more nucleic acid molecules as defined herein.

Given that we have identified a group of important proteins, such proteins are potential targets for anti-microbial therapy. It is necessary, however, to determine
10 whether each individual protein is essential for the organism's viability. Thus, the present invention also provides a method of determining whether a protein or polypeptide as described herein represents a potential anti-microbial target which comprises antagonising, inhibiting or otherwise interfering with the function or expression of said protein and determining whether *S.pneumoniae* is still viable.

15 A suitable method for inactivating the protein is to effect selected gene knockouts, ie prevent expression of the protein and determine whether this results in a lethal change. Suitable methods for carrying out such gene knockouts are described in Li
20 *et al* , *P.N.A.S.*, 94:13251-13256 (1997) and Kolkman *et al* , 178:3736-3741 (1996).

In a final aspect the present invention provides the use of an agent capable of antagonising, inhibiting or otherwise interfering with the function or expression of a protein or polypeptide of the invention in the manufacture of a medicament for use in
25 the treatment or prophylaxis of *S.pneumoniae* infection.

As mentioned above, we have used a bacterial expression system as a means of identifying those proteins which are surface associated, secreted or exported and

thus, would find use as antigens.

The information necessary for the secretion/export of proteins has been extensively studied in bacteria. In the majority of cases, protein export requires a signal peptide to be present at the N-terminus of the precursor protein so that it becomes directed to the translocation machinery on the cytoplasmic membrane. During or after translocation, the signal peptide is removed by a membrane associated signal peptidase. Ultimately the localization of the protein (i.e. whether it be secreted, an integral membrane protein or attached to the cell wall) is determined by sequences other than the leader peptide itself.

We are specifically interested in surface located or exported proteins as these are likely to be antigens for use in vaccines, as diagnostic reagents or as targets for therapy with novel chemical entities. We have therefore developed a screening vector-system in *Lactococcus lactis* that permits genes encoding exported proteins to be identified and isolated. We provide below a representative example showing how given novel surface associated proteins from *Streptococcus pneumoniae* have been identified and characterized. The screening vector incorporates the staphylococcal nuclease gene *nuc* lacking its own export signal as a secretion reporter.

Staphylococcal nuclease is a naturally secreted heat-stable, monomeric enzyme which has been efficiently expressed and secreted in a range of Gram positive bacteria (Shortle, *Gene*, 22:181-189 (1983); Kovacevic *et al.*, *J. Bacteriol.*, 162:521-528 (1985); Miller *et al.*, *J. Bacteriol.*, 169:3508-3514 (1987); Liebl *et al.*, *J. Bacteriol.*, 174:1854-1861 (1992); Le Loir *et al.*, *J. Bacteriol.*, 176:5135-5139 (1994); Poquet *et al.*, *J. Bacteriol.*, 180:1904-1912 (1998)).

Recently, Poquet *et al.* ((1998), *supra*) have described a screening vector incorporating the *nuc* gene lacking its own signal leader as a reporter to identify exported proteins in Gram positive bacteria, and have applied it to *L. lactis*. This

vector (pFUN) contains the pAM β 1 replicon which functions in a broad host range of Gram-positive bacteria in addition to the ColE1 replicon that promotes replication in *Escherichia coli* and certain other Gram negative bacteria. Unique cloning sites present in the vector can be used to generate transcriptional and translational fusions between cloned genomic DNA fragments and the open reading frame of the truncated *nuc* gene devoid of its own signal secretion leader. The *nuc* gene makes an ideal reporter gene because the secretion of nuclease can readily be detected using a simple and sensitive plate test: Recombinant colonies secreting the nuclease develop a pink halo whereas control colonies remain white (Shortle, (1983), *supra*; Le Loir *et al.*, (1994), *supra*).

Thus, the invention will now be described with reference to the following representative example, which provides details of how the proteins, polypeptides and nucleic acid sequences described herein identified as antigenic targets.

We describe herein the construction of three reporter vectors and their use in *L. lactis* to identify and isolate genomic DNA fragments from *Streptococcus pneumoniae* encoding secreted or surface associated proteins.

The invention will now be described with reference to the examples, which should not be construed as in any way limiting the invention. The examples refer to the figures in which:

Figure 1: shows the results of a number of DNA vaccine trials; and

Figure 2: shows the results of further DNA vaccine trials.

EXAMPLE 1

(i) Construction of the pTREP1-nuc series of reporter vectors

(a) Construction of expression plasmid pTREP1

The pTREP1 plasmid is a high-copy number (40-80 per cell) theta-replicating gram positive plasmid, which is a derivative of the pTREX plasmid which is itself a derivative of the previously published pIL253 plasmid. pIL253 incorporates the broad Gram-positive host range replicon of pAM β 1 (Simon and Chopin, *Biochimie*, 70:559-567 (1988)) and is non-mobilisable by the *L. lactis* sex-factor. pIL253 also lacks the *tra* function which is necessary for transfer or efficient mobilisation by conjugative parent plasmids exemplified by pIL501. The Enterococcal pAM β 1 replicon has previously been transferred to various species including *Streptococcus*, *Lactobacillus* and *Bacillus* species as well as *Clostridium acetobutylicum*, (Oultram and Klaenhammer, *FEMS Microbiological Letters*, 27:129-134 (1985); Gibson *et al.*, (1979); LeBlanc *et al.*, *Proceedings of the National Academy of Science USA*, 75:3484-3487 (1978)) indicating the potential broad host range utility. The pTREP1 plasmid represents a constitutive transcription vector.

The pTREX vector was constructed as follows. An artificial DNA fragment containing a putative RNA stabilising sequence, a translation initiation region (TIR), a multiple cloning site for insertion of the target genes and a transcription terminator was created by annealing 2 complementary oligonucleotides and extending with Tfl DNA polymerase. The sense and anti-sense oligonucleotides contained the recognition sites for NheI and BamHI at their 5' ends respectively to facilitate cloning. This fragment was cloned between the XbaI and BamHI sites in pUC19NT7, a derivative of pUC19 which contains the T7 expression cassette from pLET1 (Wells *et al.*, *J. Appl. Bacteriol.*, 74:629-636 (1993)) cloned between the EcoRI and HindIII sites. The resulting construct was designated pUCLEX. The complete expression cassette of pUCLEX was then removed by cutting with HindIII and blunting followed by cutting with EcoRI before cloning into EcoRI and SacI (blunted) sites of pIL253 to generate the vector pTREX (Wells and Schofield, *In*

Current advances in metabolism, genetics and applications-NATO ASI Series, H 98:37-62 (1996)). The putative RNA stabilising sequence and TIR are derived from the *Escherichia coli* T7 bacteriophage sequence and modified at one nucleotide position to enhance the complementarity of the Shine Dalgarno (SD) motif to the
5 ribosomal 16s RNA of *Lactococcus lactis* (Schofield *et al.* pers. coms. University of Cambridge Dept. Pathology.).

A *Lactococcus lactis* MG1363 chromosomal DNA fragment exhibiting promoter activity which was subsequently designated P7 was cloned between the EcoRI and
10 BglII sites present in the expression cassette, creating pTREX7. This active promoter region had been previously isolated using the promoter probe vector pSB292 (Waterfield *et al.*, *Gene*, 165:9-15 (1995)). The promoter fragment was amplified by PCR using the Vent DNA polymerase according to the manufacturer.

15 The pTREP1 vector was then constructed as follows. An artificial DNA fragment which included a transcription terminator, the forward pUC sequencing primer, a promoter multiple -cloning site region and a universal translation stop sequence was created by annealing two overlapping partially complementary synthetic
oligonucleotides together and extending with sequenase according to manufacturers
20 instructions. The sense and anti-sense (pTREP_F and pTREP_R) oligonucleotides contained the recognition sites for EcoRV and BamHI at their 5' ends respectively to facilitate cloning into pTREX7. The transcription terminator was that of the *Bacillus penicillinase* gene, which has been shown to be effective in *Lactococcus* (Jos *et al.*, *Applied and Environmental Microbiology*, 50:540-542 (1985)). This was considered
25 necessary as expression of target genes in the pTREX vectors was observed to be leaky and is thought to be the result of cryptic promoter activity in the origin region (Schofield *et al.* pers. coms. University of Cambridge Dept. Pathology.). The forward pUC primer sequencing was included to enable direct sequencing of cloned DNA fragments. The translation stop sequence which encodes a stop codon in 3

different frames was included to prevent translational fusions between vector genes and cloned DNA fragments. The pTREX7 vector was first digested with EcoRI and blunted using the 5' - 3' polymerase activity of T4 DNA polymerase (NEB) according to manufacturer's instructions. The EcoRI digested and blunt ended pTREX7 vector was then digested with Bgl II thus removing the P7 promoter. The artificial DNA fragment derived from the annealed synthetic oligonucleotides was then digested with EcoRV and Bam HI and cloned into the EcoRI(blunted)-Bgl II digested pTREX7 vector to generate pTREP. A *Lactococcus lactis* MG1363 chromosomal promoter designated P1 was then cloned between the EcoRI and BglII sites present in the pTREP expression cassette forming pTREP1. This promoter was also isolated using the promoter probe vector pSB292 and characterised by Waterfield *et al.*, (1995), *supra*. The P1 promoter fragment was originally amplified by PCR using vent DNA polymerase according to manufacturers instructions and cloned into the pTREX as an EcoRI-BglII DNA fragment. The EcoRI-BglII P1 promoter containing fragment was removed from pTREP1 by restriction enzyme digestion and used for cloning into pTREP (Schofield *et al.* pers. coms. University of Cambridge, Dept. Pathology.).

(b) PCR amplification of the *S. aureus* nuc gene.

The nucleotide sequence of the *S. aureus* nuc gene (EMBL database accession number V01281) was used to design synthetic oligonucleotide primers for PCR amplification. The primers were designed to amplify the mature form of the nuc gene designated nucA which is generated by proteolytic cleavage of the N-terminal 19 to 21 amino acids of the secreted propeptide designated Snase B (Shortle, (1983), *supra*). Three sense primers (nucS1, nucS2 and nucS3, Appendix 1) were designed, each one having a blunt-ended restriction endonuclease cleavage site for EcoRV or SmaI in a different reading frame with respect to the nuc gene. Additionally BglII and BamHI were incorporated at the 5' ends of the sense and anti-sense primers

respectively to facilitate cloning into BamHI and BglII cut pTREP1. The sequences of all the primers are given in Appendix 1. Three nuc gene DNA fragments encoding the mature form of the nuclease gene (NucA) were amplified by PCR using each of the sense primers combined with the anti-sense primer described above. The
5 nuc gene fragments were amplified by PCR using *S. aureus* genomic DNA template, Vent DNA Polymerase (NEB) and the conditions recommended by the manufacturer. An initial denaturation step at 93 °C for 2 min was followed by 30 cycles of denaturation at 93 °C for 45 sec, annealing at 50 °C for 45 seconds, and extension at 73 °C for 1 minute and then a final 5 min extension step at 73 °C. The
10 PCR amplified products were purified using a Wizard clean up column (Promega) to remove unincorporated nucleotides and primers.

(c) Construction of the pTREP1-nuc vectors

15 The purified nuc gene fragments described in section b were digested with Bgl II and BamHI using standard conditions and ligated to BamHI and BglII cut and dephosphorylated pTREP1 to generate the pTREP1-nuc1, pTREP1-nuc2 and pTREP1-nuc3 series of reporter vectors. General molecular biology techniques were carried out using the reagents and buffer supplied by the manufacture or using
20 standard conditions(Sambrook and Maniatis, (1989), *supra*). In each of the pTREP1-nuc vectors the expression cassette comprises a transcription terminator, lactococcal promoter P1, unique cloning sites (BglII, EcoRV or SmaI) followed by the mature form of the nuc gene and a second transcription terminator. Note that the sequences required for translation and secretion of the nuc gene were deliberately excluded in
25 this construction. Such elements can only be provided by appropriately digested foreign DNA fragments (representing the target bacterium) which can be cloned into the unique restriction sites present immediately upstream of the *nuc* gene.

In possessing a promoter, the pTREP1-nuc vectors differ from the pFUN vector described by Poquet *et al.* (1998), *supra*, which was used to identify *L. lactis* exported proteins by screening directly for Nuc activity directly in *L. lactis*. As the pFUN vector does not contain a promoter upstream of the *nuc* open reading frame the cloned genomic DNA fragment must also provide the signals for transcription in addition to those elements required for translation initiation and secretion of Nuc. This limitation may prevent the isolation of genes that are distant from a promoter for example genes which are within polycistronic operons. Additionally there can be no guarantee that promoters derived from other species of bacteria will be recognised and functional in *L. lactis*. Certain promoters may be under stringent regulation in the natural host but not in *L. lactis*. In contrast, the presence of the P1 promoter in the pTREP1-nuc series of vectors ensures that promoterless DNA fragments (or DNA fragments containing promoter sequences not active in *L. lactis*) will still be transcribed.

(d) Screening for secreted proteins in *S. pneumoniae*

Genomic DNA isolated from *S. pneumoniae* was digested with the restriction enzyme Tru9I. This enzyme which recognises the sequence 5'- TTAA -3' was used because it cuts A/T rich genomes efficiently and can generate random genomic DNA fragments within the preferred size range (usually averaging 0.5 - 1.0 kb). This size range was preferred because there is an increased probability that the P1 promoter can be utilised to transcribe a novel gene sequence. However, the P1 promoter may not be necessary in all cases as it is possible that many Streptococcal promoters are recognised in *L. lactis*. DNA fragments of different size ranges were purified from partial Tru9I digests of *S. pneumoniae* genomic DNA. As the Tru 9I restriction enzyme generates staggered ends the DNA fragments had to be made blunt ended before ligation to the EcoRV or SmaI cut pTREP1-nuc vectors. This was achieved by the partial fill-in enzyme reaction using the 5'-3' polymerase

activity of Klenow enzyme. Briefly Tru9I digested DNA was dissolved in a solution (usually between 10-20 μ l in total) supplemented with T4 DNA ligase buffer (New England Biolabs; NEB) (1X) and 33 μ M of each of the required dNTPs, in this case dATP and dTTP. Klenow enzyme was added (1 unit Klenow enzyme (NEB) per μ g of DNA) and the reaction incubated at 25°C for 15 minutes. The reaction was stopped by incubating the mix at 75°C for 20 minutes. EcoRV or SmaI digested pTREP-nuc plasmid DNA was then added (usually between 200-400 ng). The mix was then supplemented with 400 units of T4 DNA ligase (NEB) and T4 DNA ligase buffer (1X) and incubated overnight at 16°C. The ligation mix was precipitated directly in 100% Ethanol and 1/10 volume of 3M sodium acetate (pH 5.2) and used to transform *L. lactis* MG1363 (Gasson, 1983). Alternatively, the gene cloning site of the pTREP-nuc vectors also contains a BglII site which can be used to clone for example Sau3AI digested genomic DNA fragments.

L. lactis transformant colonies were grown on brain heart infusion agar and nuclease secreting (Nuc⁺) clones were detected by a toluidine blue-DNA-agar overlay (0.05 M Tris pH 9.0, 10 g of agar per litre, 10 g of NaCl per liter, 0.1 mM CaCl₂, 0.03 % wt/vol. salmon sperm DNA and 90 mg of Toluidine blue O dye) essentially as described by Shortle, 1983, *supra* and Le Loir *et al.*, 1994, *supra*). The plates were then incubated at 37°C for up to 2 hours. Nuclease secreting clones develop an easily identifiable pink halo. Plasmid DNA was isolated from Nuc⁺ recombinant *L. lactis* clones and DNA inserts were sequenced on one strand using the NucSeq sequencing primer described in Appendix 1, which sequences directly through the DNA insert.

25 Isolation of Genes Encoding Exported Proteins from *S. pneumoniae*

A large number of gene sequences putatively encoding exported proteins in *S. pneumoniae* have been identified using the nuclease screening system. These have

now been further analysed to remove artefacts. The sequences identified using the screening system have been analysed using a number of parameters.

1. All putative surface proteins were analysed for leader/signal peptide sequences using the software programs Sequencher (Gene Codes Corporation) and DNA Strider (Marck, *Nucleic Acids Res.*, 16:1829-1836 (1988)). Bacterial signal peptide sequences share a common design. They are characterised by a short positively charged N-terminus (N region) immediately preceding a stretch of hydrophobic residues (central portion-h region) followed by a more polar C-terminal portion which contains the cleavage site (c-region). Computer software is available which allows hydropathy profiling of putative proteins and which can readily identify the very distinctive hydrophobic portion (h-region) typical of leader peptide sequences. In addition, the sequences were checked for the presence of or absence of a potential ribosomal binding site (Shine-Dalgarno motif) required for translation initiation of the putative nuc reporter fusion protein.

2. All putative surface protein sequences were also matched with all of the protein/DNA sequences using the publicly databases [OWL-proteins inclusive of SwissProt and GenBank translations]. This allows us to identify sequences similar to known genes or homologues of genes for which some function has been ascribed. Hence it has been possible to predict a function for some of the genes identified using the LEEP system and to unequivocally establish that the system can be used to identify and isolate gene sequences of surface associated proteins. We should also be able to confirm that these proteins are indeed surface related and not artifacts. The LEEP system has been used to identify novel gene targets for vaccine and therapy.

3. Some of the genes identified proteins did not possess a typical leader peptide sequence and did not show homology with any DNA/protein sequences in the database. Indeed these proteins may indicate the primary advantage of our screening method, i.e. the isolation of atypical surface-related proteins, which may

have been missed in all previously described screening protocols or approaches based on sequence homology searches.

In all cases, only partial gene sequences were initially obtained. Full length genes
5 were obtained in all cases by reference to the TIGR *S.pneumoniae* database
([www@tigr.org](http://www.tigr.org)). Thus, by matching the originally obtained partial sequences with
the database, we were able to identify the full length gene sequences. In this way, as
described herein, three groups of genes were clearly identified, ie a group of genes
encoding previously unidentified *S.pneumoniae* proteins, a second group exhibiting
10 some homology with known proteins from a variety of sources and a third group
which encoded known *S.pneumoniae* proteins, which were, however, not known as
antigens.

Example 2: Vaccine trials

15

pcDNA3.1+ as a DNA vaccine vector

pcDNA3.1+

20 The vector chosen for use as a DNA vaccine vector was pcDNA3.1 (Invitrogen)
(actually pcDNA3.1+, the forward orientation was used in all cases but may be
referred to as pcDNA3.1 here on). This vector has been widely and successfully
employed as a host vector to test vaccine candidate genes to give protection against
pathogens in the literature (Zhang, *et al.*, Kurar and Splitter, Anderson *et al.*). The
25 vector was designed for high-level stable and non-replicative transient expression in
mammalian cells. pcDNA3.1 contains the ColE1 origin of replication which allows
convenient high-copy number replication and growth in *E. coli*. This in turn allows
rapid and efficient cloning and testing of many genes. The pcDNA3.1 vector has a
large number of cloning sites and also contains the gene encoding ampicillin
30 resistance to aid in cloning selection and the human cytomegalovirus (CMV)
immediate-early promoter/enhancer which permits efficient, high-level expression of
the recombinant protein. The CMV promoter is a strong viral promoter in a wide
range of cell types including both muscle and immune (antigen presenting) cells.
This is important for optimal immune response as it remains unknown as to which
35 cells types are most important in generating a protective response *in vivo*. A T7
promoter upstream of the multiple cloning site affords efficient expression of the

modified insert of interest and which allows *in vitro* transcription of a cloned gene in the sense orientation.

5 Zhang, D., Yang, X., Berry, J. Shen, C., McClarty, G. and Brunham, R.C. (1997) "DNA vaccination with the major outer-membrane protein genes induces acquired immunity to *Chlamydia trachomatis* (mouse pneumonitis) infection". *Infection and Immunity*, 176, 1035-40.

10 Kurar, E. and Splitter, G.A. (1997) "Nucleic acid vaccination of *Brucella abortus* ribosomal L7/L12 gene elicits immune response". *Vaccine*, 15, 1851-57.

15 Anderson, R., Gao, X.-M., Papakonstantinou, A., Roberts, M. and Dougan, G. (1996) "Immune response in mice following immunisation with DNA encoding fragment C of tetanus toxin". *Infection and Immunity*, 64, 3168-3173.

Preparation of DNA vaccines

20 Oligonucleotide primers were designed for each individual gene of interest derived using the LEEP system. Each gene was examined thoroughly, and where possible, primers were designed such that they targeted that portion of the gene thought to encode only the mature portion of the gene protein. It was hoped that expressing those sequences that encode only the mature portion of a target gene protein, would facilitate its correct folding when expressed in mammalian cells. For example, in the majority of cases primers were designed such that putative N-terminal signal peptide sequences would not be included in the final amplification product to be cloned into the pcDNA3.1 expression vector. The signal peptide directs the polypeptide precursor to the cell membrane via the protein export pathway where it is normally cleaved off by signal peptidase I (or signal peptidase II if a lipoprotein). Hence the signal peptide does not make up any part of the mature protein whether it be displayed on the surface of the bacteria surface or secreted. Where an N-terminal leader peptide sequence was not immediately obvious, primers were designed to target the whole of the gene sequence for cloning and ultimately, expression in pcDNA3.1.

35 Having said that, however, other additional features of proteins may also affect the expression and presentation of a soluble protein. DNA sequences encoding such features in the genes encoding the proteins of interest were excluded during the design of oligonucleotides. These features included:

- 40
1. LPXTG cell wall anchoring motifs.
 2. LXXC lipoprotein attachment sites.
 3. Hydrophobic C-terminal domain.

4. Where no N-terminal signal peptide or LXXC was present the start codon was excluded.

5. Where no hydrophobic C-terminal domain or LPXTG motif was present the stop codon was removed.

5

Appropriate PCR primers were designed for each gene of interest and any and all of the regions encoding the above features was removed from the gene when designing these primers. The primers were designed with the appropriate enzyme restriction site followed by a conserved Kozak nucleotide sequence (in most cases (NB except in occasional instances for example ID59) GCCACC was used. The Kozak sequence facilitates the recognition of initiator sequences by eukaryotic ribosomes) and an ATG start codon upstream of the insert of the gene of interest. For example the forward primer using a BamH1 site the primer would begin GCGGGATCCGCCACCATG followed by a small section of the 5' end of the gene of interest. The reverse primer was designed to be compatible with the forward primer and with a Not1 restriction site at the 5' end in most cases (this site is TTGCGGCCGC) (NB except in occasional instances for example ID59 where a Xho1 site was used instead of Not1).

20

PCR primers

The following PCR primers were designed and used to amplify the truncated genes of interest.

25

ID5

Forward Primer 5'

CGGATCCGCCACCATGGGTCTAATTGAAGACTTAAAAAATCAA 3'

Reverse Primer 5' TTGCGGCCGCCAATGCTAGACTAAACACAAGACTCA 3'

30

ID59

Forward Primer 5' CGCGGATCCATGAAAAAATCTATTCAATTTTATAGCA 3'

Reverse Primer 5' CCCTCGAGGGCTACTTCCGATACATTTTAAACTGTAGG 3'

35

ID51

40

Forward Primer 5' CGGATCCGCCACCATGAGTCATGTCGCTGCAAATG 3'

Reverse Primer 5' TTGCGGCCGCATACCAAACGCTGACATCTACG 3'

ID29

Forward Primer 5' CGGATCCGCCACCATGCAAAAAGAGCGGTATGGTTATG
3'

5 Reverse Primer 5' TTGCGGCCGCACCCCCATTCTTAATCCCTT 3'

ID50

Forward Primer 5'
10 CGGATCCGCCACCATGGAGGTATGTGAAATGTCACGTAAA 3'
Reverse Primer 5' TTGCGGCCGCTTTTACAAAGTCAAGCAAAGCC 3'

Cloning

15 The insert along with the flanking features described above was amplified using PCR
against a template of genomic DNA isolated from type 4 *S. pneumoniae* strain 11886
obtained from the National Collection of Type Cultures. The PCR product was cut
with the appropriate restriction enzymes and cloned in to the multiple cloning site of
20 pcDNA3.1 using conventional molecular biological techniques. Suitably mapped
clones of the genes of interested were cultured and the plasmids isolated on a large
scale (> 1.5 mg) using Plasmid Mega Kits (Qiagen). Successful cloning and
maintenance of genes was confirmed by restriction mapping and sequencing ~ 700
base pairs through the 5' cloning junction of each large scale preparation of each
construct.

25

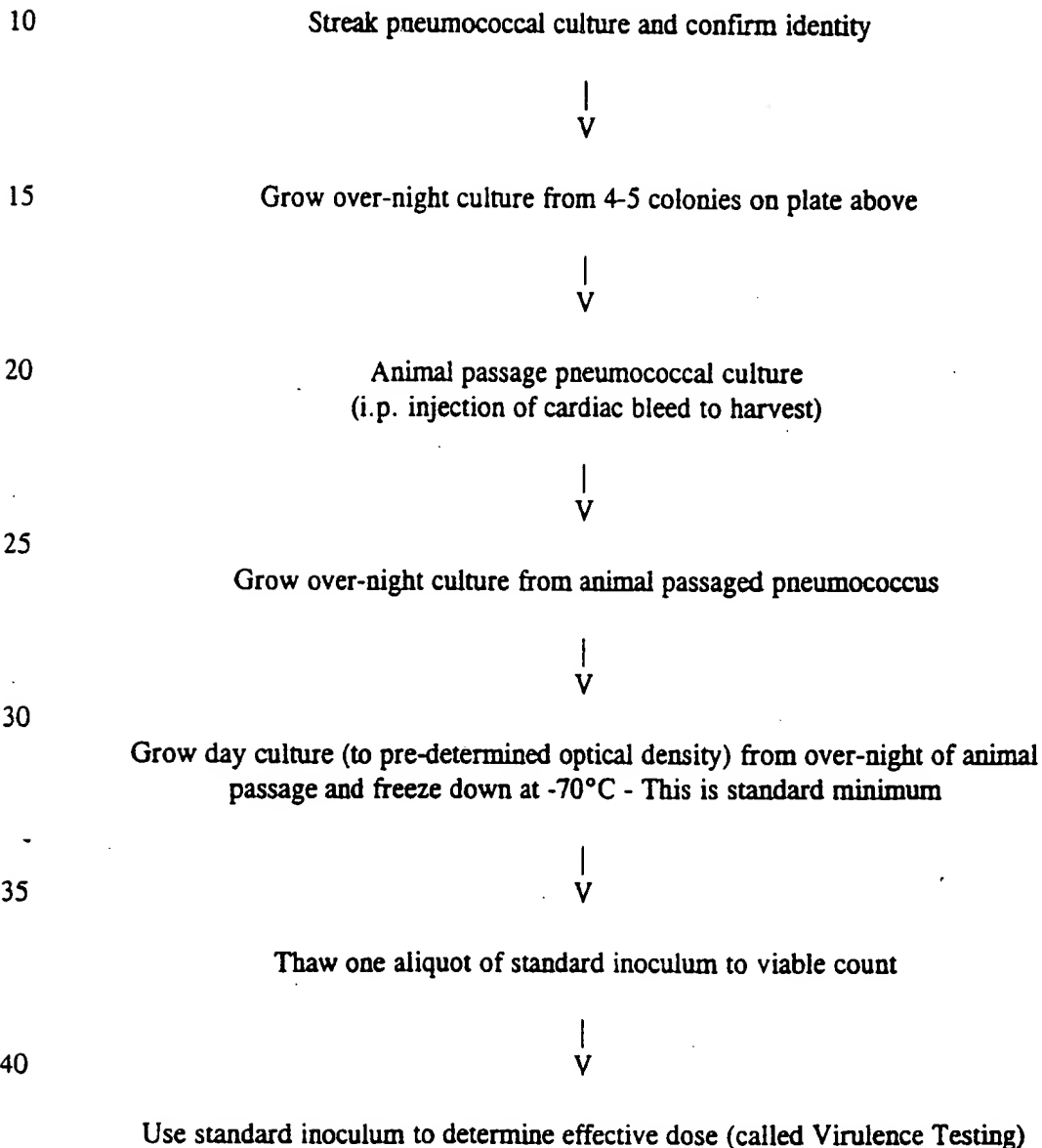
Strain validation

A strain of type 4 was used in cloning and challenge methods which is the strain
from which the *S. pneumoniae* genome was sequenced. A freeze dried ampoule of a
30 homogeneous laboratory strain of type 4 *S. pneumoniae* strain NCTC 11886 was
obtained from the National Collection of Type Strains. The ampoule was opened and
the cultured re suspended with 0.5 ml of tryptic soy broth (0.5% glucose, 5%
blood). The suspension was subcultured into 10 ml tryptic soy broth (0.5% glucose,
5% blood) and incubated statically overnight at 37°C. This culture was streaked on
35 to 5% blood agar plates to check for contaminants and confirm viability and on to
blood agar slopes and the rest of the culture was used to make 20% glycerol stocks.
The slopes were sent to the Public Health Laboratory Service where the type 4
serotype was confirmed.

40 A glycerol stock of NCTC 11886 was streaked on a 5% blood agar plate and
incubated overnight in a CO2 gas jar at 37°C. Fresh streaks were made and optochin
sensitivity was confirmed.

Pneumococcal challenge

5 A standard inoculum of type 4 *S. pneumoniae* was prepared and frozen down by passing a culture of pneumococcus 1x through mice, harvesting from the blood of infected animals, and grown up to a predetermined viable count of around 10^9 cfu/ml in broth before freezing down. The preparation is set out below as per the flow chart.



|
V

5 All subsequent challenges - use standard inoculum to effective dose

An aliquot of standard inoculum was diluted 500x in PBS and used to inoculate the mice.

10 Mice were lightly anaesthetised using halothane and then a dose of 1.4×10^5 cfu of pneumococcus was applied to the nose of each mouse. The uptake was facilitated by the normal breathing of the mouse, which was left to recover on its back.

S. pneumoniae Vaccine trials

15 Vaccine trials in mice were carried out by the administration of DNA to 6 week old CBA/ca mice (Harlan, UK). Mice to be vaccinated were divided into groups of six and each group was immunised with recombinant pcDNA3.1 + plasmid DNA containing a specific target-gene sequence of interest. A total of 100 μ g of DNA in
20 Dulbecco's PBS (Sigma) was injected intramuscularly into the tibialis anterior muscle of both legs (50 μ l in each leg). A boost was carried using the same procedure 4 weeks later. For comparison, control groups were included in all vaccine trials. These control groups were either unvaccinated animals or those administered with non-recombinant pcDNA3.1 + DNA (sham vaccinated) only,
25 using the same time course described above. 3 weeks after the second immunisation, all mice groups were challenged intra-nasally with a lethal dose of *S. pneumoniae* serotype 4 (strain NCTC 11886). The number of bacteria administered was monitored by plating serial dilutions of the inoculum on 5% blood agar plates. A
30 problem with intranasal immunisations is that in some mice the inoculum bubbles out of the nostrils, this has been noted in results table and taken account of in calculations. A less obvious problem is that a certain amount of the inoculum for each mouse may be swallowed. It is assumed that this amount will be the same for each mouse and will average out over the course of inoculations. However, the
35 sample sizes that have been used are small and this problem may have significant effects in some experiments. All mice remaining after the challenge were killed 3 or 4 days after infection. During the infection process, challenged mice were monitored for the development of symptoms associated with the onset of *S. pneumoniae* induced-disease. Typical symptoms in an appropriate order included piloerection, an increasingly hunched posture, discharge from eyes, increased lethargy and
40 reluctance to move. The latter symptoms usually coincided with the development of a moribund state at which stage the mice were culled to prevent further suffering. These mice were deemed to be very close to death, and the time of culling was used

to determine a survival time for statistical analysis. Where mice were found dead, the survival time was taken as the last time point when the mouse was monitored alive.

5 Interpretation of Results

A positive result was taken as any DNA sequence that was cloned and used in challenge experiments as described above which gave protection against that challenge. Protection was taken as those DNA sequences that gave statistically significant protection (to a 95% confidence level ($p < 0.05$)) and also those which were marginal or close to significant using Mann-Whitney or which show some protective features for example there were one or more outlying mice or because the time to the first death was prolonged. It is acceptable to allow marginal or non-significant results to be considered as potential positives when it is considered that the clarity of some of the results may be clouded by the problems associated with the administration of intranasal infections.

Results

Trials 1-6 (see figure 1)

Mouse number	Mean survival times (hours)								
	Unvacc control (1)	pcDNA 3.1+ (1)	ID5 (1)	Unvacc control (2)	ID59 (2)	Unvacc control (5)	ID59 (5)	Unvacc control (6)	ID51 (6)
1	47.5	61.0	61.0	49.0	55.0	58.0	55.3	71.6*	50.0
2	57.0	47.5	61.0	51.0	55.0	75.0	98.0	60.7	99.9T
3	47.5	50.5	57.0	49.0	55.0	48.0	58.5	98.5	53.6
4	47.5	50.5	72.0	55.0	69.5	46.7	55.3	(101.2)*T	99.9
5	77.0	72.0	47.5	49.0	74.0	58.0	53.5	60.7	59.4
6	57.0	50.5	mouse died	49.0	mouse died	75.0	98.0	50.8	50.0*
Mean	55.6	55.3	59.7	50.3	61.7	60.1	69.8	68.4	68.8
sd	11.5	9.4	8.8	2.4	9.3	12.5	21.9	18.3	24.4
p value 1	-	-	0.1722	-	0.0064	-	0.2862	-	<36.0
p value 2	-	-	0.2565	-	-	-	-	-	-

* - bubbled when dosed so may not have received full inoculum.

T - terminated at end of experiment having no symptoms of infection.

Numbers in brackets - survival times disregarded assuming incomplete dosing

p value 1 refers to significance tests compared to unvaccinated controls

p value 2 refers to significance tests compared to pcDNA3.1 + vaccinated controls

Statistical Analyses.

5 Trial 1 - None of the other groups had significantly longer survival times than the controls. The survival times of the unvaccinated and pcDNA3.1 control groups were not significantly different. One of the mice from ID5 was an outlying result and the mean survival times for ID5 were extended but not significantly so.

Trial 2 - The group vaccinated with ID59 had significantly longer survival times than the unvaccinated control group.

10 Trial 5 - The group vaccinated with ID59 again survived for an average of almost 10 hours longer than the controls but the results were not quite statistically significant.

Trial 6 - The group vaccinated with ID51 did not have survival times significantly higher than unvaccinated controls ($p = < 36.0$), however, there were 2 outlying mice in the vaccinated group.

15

Vaccine trials 7 and 8 (See figure 2)

Mouse number	Mean survival times (hours)			
	Unvacc control (7)	ID29 (7)	Unvacc control (8)	ID50 (8)
1	59.6	73.1	45.1	60.6
2	47.2	54.8	50.8	60.6
3	59.6	59.3	60.4	51.1
4	70.9	54.8*	55.2	60.6
5	68.6*	59.3	45.1	60.6
6	76.0	54.8	45.1	60.6
Mean	63.6	59.35	50.2	59.1
sd	10.3	7.1	6.4	3.9
p value 1	-	<39.0	-	0.0048

* - bubbled when dosed so may not have received full inoculum.

20 T - terminated at end of experiment having no symptoms of infection.

Numbers in brackets - survival times disregarded assuming incomplete dosing
p value 1 refers to significance tests compared to unvaccinated controls

Statistical Analyses.

- 5 Trial 7 - The ID29 vaccinated group showed prolonged times to the first death. T
Trial 8 - The group vaccinated with ID50 survived significantly longer than
unvaccinated controls.

Appendix I - Oligonucleotide primers

nucS1

Bgl II Eco RV

5 5'- cgagatctgatatctcacaaacagataacggcgtaaataag -3'

nucS2

Bgl II Sma I

10 5'- gaagatcttccccgggatcacaaacagataacggcgtaaataag -3'

nucS3

Bgl II Eco RV

5'- cgagatctgatatccatcacaaacagataacggcgtaaataag -3'

15 nucR

Bam HI

5'- cgggatccttatggacctaatacagcgttgtc -3'

NucSeq

20 5'- ggatgcittgttcaggtgtatc -3'

pTREPF

5'- catgatatcggtacctaagctcataatgtccggcaatggtgtgggctttttgttttagcggataa
caatttcacac -3'

25

pTREPR

5'- gcggatccccgggcttaattaatgttaaactagtcgaagatctcgcgaattctcctgtgtgaaatt
gttatccgcta -3'

30

pUCF

5'- cgccagggttttccagtcacgac -3'

VR

5'- tcagggggcgaggacctatg -3'

35

V1

5'- tcgtatgttgttggaattgtg -3'

V₂

5'- tccggctcgtatgtgtgtggaattg -3'

TABLE 1

5	<u>ID4 1200 bp</u>
10	<p>ATGAGAAATATGTGGGTGTAATCAAGGAAACCTATCTTCGACATGTGAGTCATGGAGTTCTTCTTTATGGTGA TTTCGCCGTTCTCTTTTAGGAATCTCTGTAGGAATTGGGCATCTCCAAGGTTCTTCTATGGCTAAAAATAATAA AGTGGCAGTAGTGACAAACAGTGCCATCTGTAGCAGAAGGACTGAAGAAATGTAATGGTGTTAACTTCGACTATAA AGACGAAGCAAGTGCCAAAGAAGCAATTAAGAAAGAAAAATTAAGAGTTATTTGACCATTGATCAAGAAAGATA GTGTTCTAAAGGCAGTTTATCATGGCGAAACATCGCTTGAAAATGGAATTAATTTGAGGTTACAGGTACACTCA ATGAACTGCAAAATCAGCTTAATCGTTCAACTGCTTCTGTCTCAAGAGCAGGAAAAACGCTTAGCGCAGACAA TTCAATTCACAGAAAAGATTGATGAAGCCAAAGGAA/ATAAAAAGTTTATTCAAACAATTGCAGCAGGTGCCTTAG GATTTCTTTCTTTATATGATTCTGATTACCTATGCGGGTGTAACAGCTCAGGAAGTTGCCAGTGAAAAAGGCCACAA AATTATGGAAGTCGTTTTTCTAGCATAAGGGCAAGTCACTATTTCTATGCGCGGATGATGGCTCTGTTCTAGTG 15 ATTGAAACGCATATTGGGATCTATGTTGTAGGTGGTCTGGCTGCCGTTTGTCTTTAAAGATTGCCATTCTTGGC TCAGTCTGGTATTTGGATCACTTGGGAGATGCTATCTCACTGAATACCTTGCTCTTTATTTTGATCAGTCTTTCA TGTACGTAGTCTTGGCAGCCTTCTAGGATCTATGGTTTCTCGTCTGAGGACTCAGGAAAGCCTTGTCGCTTT GATGATTTTGATTATGGGTGGTTTTTTGGAGTGACAGCTCTAGGTGCAGCTGGTGACAACTCTCTTGAAGATT 20 GGTCTTATATCCCTTATTTTCGACCTTCTTTATGCCGTTTTCGAACGATTAATGACTATGCGGGGGAGCAGAAG CATGGATTTCACCTTGCTATTACAGTGATTTTGGCGTGGTAGCAACAGGATTTATCGGACGCATGTATGCTAGTCT CGTTCTTCAAACGGATGATTTAGGGATTGGAACCTTTAAACGTGCCTTATCTTATAAATAG</p>
25	<p>MRNMWVVIKETYLRHVESWSFFFMVISPFLFLGISVGIGHLQSSMAKNNKVAVVTVPSVAEGLKNVNGVNFYKD EASAKEAIKEEKLKGYLTIDQEDSVLKAVYHGETSLENGIKFEVTGTLNELQNLRSTASLSQEQEKRLAQTIQFTEKI DEAKENKKFIQTIAAGALGFFLYMILITYAGVTAQEVASEKGTKIMEVVFSSIRASHYFYARMMALFLVILTHIGYVVG GLAAVLLFKDLPFLAQSGILDHLGDAISLNTLLFILISLFMYVVLAAFLGSMVSRPEDSGKALSPLMILMGFFGVLTALG AAGDNLLLKIGSYIPFISTFFMPFRTINDYAGGAEAWISLAITVIFAVVATGFIGRMYASLVLTQDDLGWKTKRALS 30 Z</p>
30	<u>ID5 1125 bp</u>
35	<p>CCTGGGAAAGTCTTGAAAATTATGATAGAATGGTGGAAGGAAAAATTGAGGAGAGTAGTAGTGACTCAAAATGTT GAAAGTCTTCTCGTATCCATTGTAATCAGTGATACAAATGAAGAAAAATATCTGCCTGGTCTAATTGAAGACTTAA AAAATCAAACCTATCCTAAAGAGGATATTGAAATCTATTTATAAATGCTATGTCCACAGATGGGACCACAGCTA TCATTAGCAATTTATAAAGGAAGATACAGAGTTAACTCAATTAGATTGTATAACAATCCTAAGAAAAATCAAG CTAGTGGTTTTTAACCTGGGAGTTAAACATTCTGTAGGGGACCTTATTTAAAAATTGATGCTCATTCAAAAGTTAC TGAGACTTTTGTAAATGAACAATGTGGCTATTATCAACAAGGTGAATTTGTCTGTTGGGGGGCCTAGACCGACGATT GTCGAAGGAAAAAGGAAAAATGGGCAGAGACCTTGCACTTGTGTAGGAAAAATATGTTTGGCAGTAGCATTGCCAAT TATCGAAAATAGTTCTGAGGATAGATATGTTCTTCTATTTTTCATGGAATGTATAAACGAGAGGTTTTCCAGAAGG 40 TTGGTTTGTAAATGAGCAACTTGGCCGAAGTAAATGATATTCAATTATAGAATTCGAGAATATGGTTATAA AATCCGCTATAGCCCAAGTATTCTATCTTATCAGTATATTCGACCAACATTCAAGAAAAATGCTGCATCAAAAGTAT TCAAATGGTTTGTGGATTGGCTTGACAAGTCATGTTTCAAGTTAAGTGTATATCATTATTTCACTATGTTCTTGT ATTGTTTTTGTAGTCTTGTGTTTGTCTAGCATTGTTACCGATCACATTGCTATTCATAACTTTACTATTAGGTGCT 45 ATTTTCTACTTTTGTCACTTACTCACTTTGCTGACTTTATTAACCATAAAAATGGATTCTAATTGTGATGCCCTT ATTTTATTTTCCATTCACTTTGCTTATGGCCTTGGGACGATTGTAGGTTAATTAGAGGATTTAAATGGAAGAAGG AGTACAAGAGAACAATAATTTATTTGGATAAAATAAGCCAAATAAATCAAAATATGCTATAA</p>
50	<p>PGKVLKIMIEWWKEKFRRVVVTQNVESLLVSVISAYNEEKYLPGLIEDLNQTYPKEDIEILFINAMSTDGTTAIIQQFIK EDTEFNSIRL YNNPKKNQASGFNLGVKHSVGDILKIDAHSKVTETFVMNNVAIIQQGEFVCGGPRPTIVEGKGKWAET LHLVEENMFSSIAN YRNSSEDRYVVSIFHGMKREVFKVGLVNEQLGRTEENDIHYRIREYGYKIRYSPSILSYQYIRP TFKKMLHQKYSNGLWIGLTSHVQFKCLSLFHYVPCLFVLSLVFSLALLPITFVFTLLLGAYFLLSLLTLLKHKRNGF LIVMPFILFSIHFA YGLGTIVGLIRGFKWKKEYKRTIYLDKISQINQNMZ</p>
55	<u>ID11 696 bp</u>
60	<p>ATGATGAAAGAACAAAATACGATAGAAATCGATGTATTCAATTAGTTAAAAGCTTGTGGAAACGCAAGCTAATG ATTTAATAGTGGCACTTGTGACAGGTGCGGGGGCTTTTGCATATAGCACTTTTATTGTTAAGCCAGAATATACGA GTACCACGCGAATTTACGTAGTGAATCGCAATCAAGGAGACAAGCCGGGGTTGACAAATCAGGATTTGCAGGCAG GAACCTATCTGGTAAAAGACTACCGTGAGATTATCTTTTCGAGGATGTTTGGAGGAAGTTGTTCTGATTGAA ACTAGATTTGACGCCAAAAGGTTTGGCTAATAAAAATTAAGTGACAGTACCAGTTGATACCCGTATTGCTCTATT TCAGTTAATGATCGAGTTCCTGAAGAGGCAAGCCGTATCGCTAACTCTTTGAGAGAAGTAGCTGCTCAAAAAAT ATCAGTATTACTCGTGTCTGACGTGACAACACTGGAGGAGGCAAGGCCGGCGATATCCCCGTCTCGCCAAAT ATTAACGCAATACACTAATTGGTTTTTTGGCAGGGGTGATTGGAAGTATAGTTCTTCTATCTTGAACCTTTT</p>

GGATACTCGTGTGAAACGTCGGAAGATATCGAAAATACATTGCAGATGACACTTTTGGGAGTTGTGCCAAACTT
GGTAAGTTGAAATAG

5 MMKEQNTTEIDVFQLVKSLWKRKLMILVALVTGAGAFAYSTFVKPEYTSSTRJYVVRNQGDKPGLTNQDLQAGTYL
VKDYREILSQDVL EEVSDLKLDLTPKGLANKIKVTPVDTRIVSISVNDRVPEEASRIANSLREVAQKIISITRVSDVT
TLEEARPAISPSPNIKRNTLIGFLAGVIGTSVIVLHLELLDTRVTRKPEDIENTLQMTLLGVVPLNGLKLZ

ID12 555 bp

10 ATGGTAAAAGTAGCAGTTATATTAGCTCAGGGCTTTGAAGAAATTGAAGCCTTGACAGTTGTAGATGTCTTGCGTC
GAGCCAATATCACATGTGATATGGTTGGTTTTGAAGAGCAAGTAACGGGTTGCGATGCAATCCAAGTAAGAGCAG
ATCATGTCTTTGATGGAGATTTATCAGACTATGATATGATTGTTCTTCTGAGGATATGCTGTTCTGCACATTTA
CGTGATAATCAGACCTTGATTCAAGAATTGCAAGCTTCGAGCAAGAAGGGAAGAACTAGCAGCCATTTGTGCG
15 GCACCAATTGCCCTCAATCAAGCAGAGATATTGAAAAATAAGCGATACACTTGTTATGACGGCGTTCAAGAGCAA
ATCCTTGATGGTCACTACGTCAAGGAAACAGTAGTGGTAGATGGTCAGTTGACAACCAAGTCGGGGTCTTCAACA
GCCCTTGCTTTGCCTACGAGTTGGTGGAGCAACTAGGAGGGGACGCAGAGAGTTTACGAACAGGAATGCTCTAT
CGAGATGCTCTTGGTAAAAATCAGTAA

20 MVKVAVILAQGFEEIEALTVVDLRRANITCDMVGFEQVTGSHAIQVRADHVFDDGLSDYDMIVLPGMMPGSAHLR
DNQTLIQELQSFEQEGKLAACAIPALNQAEILNKRYTCYDGVQEQILDGHYVKETVVVDGQLTTSRGPSTALAF
YELVEQLGGDAESLRTGMLYRDVFGKNQZ

ID27 306 bp

25 GTGGTAGGGATGGTAGAACCAAACTAGAAAGCCTTATAAAAGATCTTTACAATCATGCTCGACATGATTTGAGT
GAAGATTAGTTGCTGCTCTCCTAGAGACTACTAAAAAAGCTCTACTACAAATGAGCAATTGCAGGCAGTTGCTG
TCTCAGGCCTGGTCAATCGTGAATTGCTCCTAAATCCCAACATCCAGCACCTGAGTTGCTCAACTTGGCTCGCTT
TGTCAAAAGAGAAGAAGCCAAGTACAGAGGAAGTCCGACTTCTGCGCTTATGTATGAGGAAGTCTTTAAAAATGCT
30 TTGA

30 MVGMVEPNLESLIKDLYNHARHDLSEDLVAALLETTKKLPTTNEQLQAVRLSGLVNRLLLLNPKHPAPELLNLARFVK
REEAKYRGATATSAALMYEELFKMLZ

ID29 945 bp

35 TTGTTCTTAAAAAAGGAAAAGAGAGGTAATCAGCATGCGTAAATGGACAAAAGGATTTCTCATCTTTGGTGTGGTG
ACTACCGTTATCGGCTTTATCCTGCTTTTTGTAGGTATCCAATCTGACGGGAATAAGAGCCTACTTTCCATGTCCAA
AGAACCTGTCTATGATAGCCGTACGGAAAAGCTAACCTTTGGCAAGGAAGTCGAAAACCTAGAAATTAATCTCCA
40 CCAACACACGCTCACCATCACAGACTCTTTGATGATCAAAATCCACATTTCTTACCATCCATCTCTTTGCTCAC
CATGATCTTATACCAATCAGAACGATAGAACTCTGAGTCTCACTGATAAGAACTGTCTGAACTCCGTTTCTCT
CTTCTGGAATTGGTGGGATTCTTCATATCGCAAGTAGCTACTCTAGTCGTTTTGAAGAAGTTATTTCTCCGACTACC
AAAAGGGAGAACTCTAAAAGGGATCAACATCTCAGCCAATCGCGGACAAAACCAATCATAAATGCTAGCCTTGA
AAATGCGACCTCAATACAAACAGCTATATCCTCGAATTGAAGGAAGTCGTATCAAAAACAGTAACTCACAAAC
45 GCCCAATATCGTTAATATCTTTGATACAGTTCTTACAGATAGTCAGTAGAGTCAACAGAGAAATCACTTCCAGCT
GAAAATATCCAAGTCCATGGCAAGGTTGAAGTCTGCTGCAAGATTATCTCAGAAATCATCCTAGACCAGAAAGAA
AGCCAACGAATTAAGTGGGACATCTCAAGCAACTATGGTTCTATCTTCCAATTCAAGAGAGAAAAGCCTGAATCA
AGAGGTACGGAATTAAGCAACCTTACAAAAGTAAAAAAGCGATGTCAAGGATCAACTCATTGCGAGATCTGAT
50 GATAATATTGATCTAATATCCACACCAAGCAGACGTTGA

50 MFLKKEREVISMRKWTGKFLIFGVVTTVIGFILLFVGIQSDGIKSLSMSKEPVYDSRTEKLTFGKEVENLEITLHQHTLTI
TDSFDDQIHISYHPSLSAHDLTINQNDRTLSTDKKLSETPFLSSGIGGILHIASSYSSRFEEVILRLPKGRTLKGINISANR
GQTTIINASLENATLNTNSYLRIEGRISRIKNSKLTTPNIVNIFDVTLDTSQLESTENHFHAENIQVHGKVELTAKDYLRIL
QKESQRINWDISSNYGSIFQFTREKPESRGTELSNPYKTEKTDVKDQLIARSDDNIDLISTPSRRZ

55 ID30 879 bp

ATGAAACAAGAATGGTTTGAAGTAATGATTTTGTAAAAACAACAAGCAAGAACAAGCCTGAAGAGCAAGCTCA
AGAGGTTGCAGACAAGGCTGAAGAAACGATAGCCGATCTCGATACACCAATTGAAAAAAATCTCAGTTAGAGG
AGGAAGTCCCTCAAGCTGAAGTCGAATTGGAAAGCCAGCAAGAAGAGAAAAATTGAAGCTCTGAAGACAGTGAA
60 GCGAGAACAGAAATAGAAGAAAAGAGGCATCTAATTCTACTGAAGAAGAGCCAGACCTTTCTAAGAAAAAGAA
AAAAGTCACTATAGCTGAAGAGAGCCAAGAGCTCTTCTCAGCAAAAAGCAACCAAGAAAGAGCCACTTCTTAT
CAGTAAATCTTTAGAAAGTCTTATATCCCCGACCAAGCTCCAAAATCTAGGGATAAATGGAAGAGCAAGTGCT
TGATTTTTGGTCTTGGCTAGTGGAAGCGATCAAAATCTCTACAAGTAAGTTGGAACAAGTATCACACACAGTTAC
ACAGCCTTTCTCTGCTCATTCTGTTTTCTGCATCTTCTTTTCTTTAGTATCTATCATCAACATGCTTACTAT
65 GGACATATAGCAAGCATTAAACAGTCGCTTCCCTGAGCAGCTAGCTCCTTAACTCTTTTTCTATCATCTCTATCCT

AGTAGCGACAACACTCTTCTTTCTTTTCATTCTCTTGGGTAGTTTCGTTGTGAGACGATTTATCCACCAGGAAAAAG
GACTGGACGCTAGACAAGGTTCTCCAACAATATAGTCAACTCTTGGCAATTCCAATCTCCTCACTGCTATTGCTAG
TTTCTTTGCTTTCTTTGATAGCCTACGATTACAGCCCTCTTGTGTGTGA

5 MKQEWESNDFVKTSSKNKPEEQAEVADKAEETIADLDTPIEKNTOLEEEVPQAEVELESQOEEKIEAPEDSEARTEIE
EKKASNSTEEEPDLSEKETEKVIAEESQEALPQQKATTKPELLSKSLESPIPDQAPKSRDKWKEQVLDWFWSWLVEAKS
PTSKLETSITHSYTAFLLLLFSASSFFSIYHIKHAYYGHASINSRFEQLAPLTLFSIISILVATTLFFFSFLLGFSVVRRIH
QEKDWTLDKVLQQYSQLLAIPISLLLLVSLLSLIA YDLQPSVZ

10 ID105 990 bp

ATGCAACTCGCTTCTTCGGTCTACTCATTGTTCTGCTGGTACAATTTGTTCTTAAAAAAGGAAAGAGAGGTAATCA
GCATGCGTAAATGGACAAAAGGATTTCATCTTTGGTGTGGTACTACCGTTATCGGCTTTATCCTGCTTTTTGTA
GGTATCCCAATCTGACGGGATTAAGAGCCTACTTTCCATGTCCAAAAGACCTGTCTATGATAGCCGTACGGAAAAAG
15 CTAACCTTTGGCAAGGAAGTCGAAAACCTAGAAATTACTCTCCACCAACACACGCTCACCATCACAGACTCTTTC
GATGATCAAATCCACATTTCTTACCATCCATCTCTTCTGCTCACCATGATCTTATCACCATCAGAACGATAGAA
CTCTGAGTCTCACTGATAAGAACTGTCTGAACTCCGTTTCTCTCTTCTGGAATTGGTGGGATTCTTCATATCGC
AAGTAGCTACTCTAGTCGTTTTGAAGAAGTTATTCTCGACTACAAAAGGGAGAACTCTAAAAGGGATCAACAT
20 CTCAGCCAATCGCGGACAAACCACCATCATAAATGCTAGCCTTGAATAATGCGACCCTCAATACAAACAGCTATAT
CCTCCGAATTGAAGGAAGTCGTATCAAAAACAGTAACTCACAACGCCCAATATCGTTAATATCTTTGATACAGTT
CTTACAGATAGTCAGCTAGAGTCAACAGAGAATCACTTCCACGCTGAAAAATCCTCAAGTCCATGGCAAGGTTGAA
CTGACTGCCAAAGATTATCTCAGAATCATCTAGACCAGAAAGAAAGCCAAACGAATTAAGTGGGACATCTCAAGC
AACTATGTTTCTATCTTCCAATTCAAGAGAAAAGCCTGAATCAAGAGGTACGGAATTAAGCAACCCCTTACAAA
25 ACTGAAAAAACCGATGTCAAGGATCAACTCATTGCGAGATCTGATGATAATATTGATCTAATATCCACACCAAGC
AGACGTTGA

30 MQLASSVYSLFVWYNLFLLKKEREVISMRKWTGFLIFGVVTTVIGFILLFVGIOQSDGIKSLLSMSKEPVYDSRTEKLTFG
KEVENLEITLHQHTLTITDSFDDQIHISYHPSLSAHHDLTINQNDRTLSLTDKKLSETPFLSSGIGGILHIASSYSSRFEEVIL
RLPKGRITLKGINISANRGQTTIHASLENATLNTSYILRIEGRIRKNSKLTTPNIVNIFDTVLTDQLESTENHFHAENIQV
HGKVELTAKDYLRILDQKESQRINWDISSNYGSIFQFTREKPESRGTELSNPYKTEKTDVKDQLIARSDDNIDLISTPSRR
Z

ID107 -78bp

35 ATGATATGTAAAAATGAAGCAGGGAGGGAGCAGGGCGTGCTGGGGATGGAGAGTGGGGGAGGGACGCTGCTATTT
TAATC

40 MICKMKQGGSRACWGWVRVGEGRCYFN

ID109 714 bp

45 CGATAAAGAGGCCTTGAGTAATCTCAATTTGCAGATTGAAAATGGAGAGATTATGGGCTTGATTGGTCATAATGG
GGCTGGAAAAATCGACCACTATAAAATCCCTAGTCAGTATCAATTCACCCAGCAGTGGTCGTATTTTGGTAGACGGT
CAGGAGTTATCGGAAAAATCGCTTGGCTATTAACGAAAAGATTGGCTACGTAGCAGACTCGCCTGACTTATTTTAC
GCTTAACGGCCAATGAATTTTGGGAATTGATCGCCTCATCTATGATCTGAGTAGATCTGACTTGGAGGCTAGTCT
AGCTAGGCTATTGAACGTTTTTGATTTGCTGAAAATCGCTATCAGGTTATTGAACTCTTCTCACGGAATGCGT
50 CAGAAAGTCTTTGTCATCGGAGCACTCTGTCTGATCCCGATATTTGGGTTTTGGACGAACCCCTTGACTGGTTTG
ATCCCAGGCTGCCTTTGATTGAAACAGATGATGAAGGAACATGCACAAAAAGGGAAGACAGTCTTGTTTTCAA
CTCATGTCCTAGAGGTGGCAGAGCAAGTCTGTGATCGGATTGCCATTTTGA AAAAGGGGCATTGATTATTGTGG
TAAGGTAGAGGACTTGAGGAAAGACCACCCAGACCAGTCTTTGAAAAGTATCTACCTTAGTCTTGCTGGTAGAAA
AGAGGAGGTTGCGGATGCGTCTCAAGGTCATTA

55 DKEALSNNLQIENGEMGLIGHNGAGKSTTIKSLVSISSPSSGRILVDGQELSENRLAIKRKIGYVADSPDLFLRLTANEF
WELIASSYDLSRSDLEASLARLLNVDFEAENRYQVIETLSHGMQRQKVFVIGALLSDPDIWVLDPLTGLDPQAAFDLQK
MMKEHAQKGTVLFSTHVLEVAEQVCDRIAILKKGHLIYCGKVEDLRKDHPDQSLSIYLSLAGRKEEVADASQGHZ

60 ID112 360 bp

ATGGCTTTGTTTTTCAGAGAGAGGAGCAGTACGGAAGACACCAATGGCAAGTCCAATAATGAGACCTATGATGGTT
CCGACGATAGAGATTAAAAGAGTGATACCAGCACCAAGAGTTGTTGCCAGTTTTAGGAAAGAAATTTAGCA
ACTTGGCTAAAGAACTACTGCTAGTCTCTTCAGTTGTTGTAGCTTCGGCAGGTTGTTCTTGTATCATACGATCCA
65 TCAAGGCAACTTGGTCATCTTTGAAATGGTTTCAATGCTGGCATTGATTGGCTAATACGATTGTCATTTTACGA
AGCCCGATAGCGATAGCTGTATCTTCTCCCAAGTTTGA AACCAAGGTTCTACTTGA

MALFSERGAVRKTPTMASPIMRPMVPTIEIKRVIPAPRKSCQFSEIRLATWLKLLLVSSVVASAGCSLIIRSIKATWSS
FEMVSMALILWLRLSFLRSPAIIVSSSPVLKPGSTZ

5 ID 128 - 3.43

ATGAAATTTAGTAAAAATATATAGCAGCTGGATCAGCTGTTATCGTATC
CTTGAGTCTATGTGCCTATGCACTAAACCAGCATCGTTCGCAGGAAAAA
10 AGGACAATAATCGTGTCTCTTATGTGGATGGCAGCCAGTCAAGTCAGAAA
AGTGAAAACCTTGACACCAGACCAGGTTAGCCAGAAAGGAATTCAGGC
TGAGCAAATTGTAATCAAAATTACAGATCAGGGCTATGTAACGTACACG
GTGACCACTATCATTACTATAATGGGAAAGTTCCTTATGATGCCCTCTTT
AGTGAAGAAGCTTTGATGAAGGATCCAACTATCAACTTAAAGACGCTGA
TATTGTCAATGAAGTCAAGGGTGGTTATATCATCAAGGTCGATGGAAAAAT
15 ATTAGTGTCTACCTGAAAAGATGCAGCTCATGCTGATAATGTTCCGAATAAA
GATGAAATCAATCGTCAAAAAAAGAACATGTCAAAGATAATGAGAAGGT
TAACTCTAATGTTGCTGTAGCAAGGCTCTCAGGGACGATATACGACAAATG
ATGGTTATGTCTTTAATCCAGCTGATATTATCGAAGATACGGGTAATGCT
TATATCGTTCCTCATGGAGGTCATATCACTACATTCCCAAAAGCGATTT
20 ATCTGCTAGTGAATTAGCAGCAGCTAAAGCACATCTGGCTGGAAAAATA
TGCAACCGAGTCAGTTAAGCTATTTCTCAACAGCTAGTGACAATAACACG
CAATCTGTAGCAAAAGGATCAACTAGCAAGCCAGCAATAAATCTGAAAA
TCTCCAGAGTCTTTTGAAGGAACTCTATGATTACCTAGCGCCCAACGTT
ACAGTGAATCAGATGGCCTGGTCTTTGACCCTGCTAAGATTATCAGTCGT
25 ACACCAATGGAGTTGCGATTCCGCATGGCGACCATTACCACTTTATTCC
TTACAGCAAGCTTTCTGCCTTAGAAGAAAAGATTGCCAGAATGGTGCCTA
TCAGTGGAACTGGTCTACAGTTTCTACAAATGCAAAACCTAATGAAGTA
GTGTCTAGTCTAGGCAGTCTTTCAAGCAATCCTTCTTTAACGACAAG
TAAGGAGCTCTCTCAGCATCTGATGGTTATATTTTAAATCCAAAAGATA
30 TCGTTGAAGAAACGGCTACAGCTTATATTGTAAGACATGGTGATCATTTC
CATTACATTCAAAAATCAAATCAAATGGGCAACCGACTCTTCCAAACAA
TAGTCTAGCAACACCTTCTCCATCTCTTCCAATCAATCCAGGAACCTCAC
ATGAGAAAACATGAAGAAGATGGATACGGATTTGATGCTAATCGTATTATC
35 GCTGAAGATGAATCAGGTTTTGTCATGAGTCACGGAGACCACAATCATT
TTTCTTCAAGAAGGACTTGACAGAAGAGCAAATTAAGGTGCGCAAAAAA
TTTAG

MKFSKKYIAAGSAVIVSLSLCAYALNQHRSQENKDNRRVSYVDGSQSSQK
40 SENLTPDQVSQKEGIAEQIVIKITDQGYVTSBGDHYHYNGKVPYDALF
SEELLMKDPNYQLKDADIVNEVKGYYIIVKVDGKYVYVYKDAHADNVRTK
DEINRQKQEHVKDNEKVNSNVAVARSQGRYTTNDGYVFNPAHIEDTGNA
YIVPHGGHYHYIPKSDLSASELAAAKAHLAKNMQPSQLSYSSTASDNNT
QSVAKGSTSKPANKSENLSLLKELYDSPAQRYSSEGLVFDPAKIISR
TPNGVAIPHGDHYHFIPYSKLSALEEKIARMVPISGTGSTVSTNAKPNEV
45 VSSLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHF
HYIPKSNQIGQPTLPNNSLATPSPSLPINPGTSHEKHEEDGYGFDANRJI
AEDESGFVMSHGDHNYFFKKDLTEEQIKVRKNI*

TABLE 2

ID2 840 bp

5 ATGGGAATTGCTCTAGAAAATGTGAATTTTACATATCAAGAAGGTAAGTCCCTTAGCTTCAGCAGCTTTGTCGGATG
 TTTCTTTGACGATTGAAGATGGCTCTTATACAGCTTTAATTGGGCACACAGGTAGTGGTAAATCAACTATTTTACA
 ACTCTTAAATGGTTTATTGGTGCCAAAGTCAAGGGAGTGTGAGGGTTTTTGATACCTTAATCACTCGACTTCTAAA
 AATAAAGATATTTCGCAAAATTAGAAAACAGGTTGGCTTGGTATTTTCAGTTTGCTGAAAATCAGATTTTGAAGAAA
 CGGTTTTGAAGGACGTGCTTTTGGACCCGCAAAATTTTGGAGTTTCTGAAGAAGATGCTGTGAAGACTGCCGCTGA
 10 GAACTGGCTCTGGTTGGAATTGATGAATCACTTTTGTATCGTAGTCCGTTTGAGCTGTGAGGGGACAAATGAGA
 CGTGTGGCCATTGCAGGCATACCTGCCATGGAGCCAGCTATATTAGTCTTAGATGAGCCACAGCTGGTCTAGATC
 CTCTAGGGAGAAAAGAGTTGATGACCCTGTTCAAAAACTCCACCAGTCAGGGATGACCATCGTCTTGGTAAACGC
 ATTTGATGGATGATGTTGCTGAATATGCGAATCAAGTCTATGTAATGGAAGGGGACGTTTGTAAAGGGGGGCA
 AACCAGTGATGCTTTCAAGACGTTGTTTTATGGAAGAAGTTCAGTTGGGAGTACCTAAAAATTACGGCCTTTTG
 15 TAAACGATTGGCTGATAGAGGCGTGTCATTTAAACGATTACCGATTAAGATAGAGGAGTTCAAGGAGTCGCTAAA
 TGGATAG

MGIALENVNTFYQEGTPLASAALSDVSLTIEDGYSYALIGHTGSGKSTILQLLNGLLVPSQGSVRVFDLTITSTSKNKDIR
 20 QIRKQVGLVFQFAENQIFETVVKDVAFGPQNFVSEEDAVKTAREKLALVGIDSLFDRSPFELSGGQMRRVVAJAGILA
 MEPAILVLDEPTAGLDPLGRKELMTLFKKLHQSGMTIVLVTHLMDDVAEYANQVYVMEKGRLVKGKPSDVFQDVV
 FMEEVQLGVPKITAFCKRLADRGVSFKRLPIKIEEFKESLNGZ

ID 3 6360 bp

25 TACCCGGTAGTCTTAGCAGACACATCTAGCTCTGAAGATGCTTTAAACATCTCTGATAAAGAAAAAGTAGCAGAA
 AATAAAGAGAAAACATGAAAATATCCATAGTGCTATGGAACCTTCACAGGATTTTAAAGAGAAAACAGCAGTC
 ATTAAGGAAAAAGAGTTGTTAGTAAAAATCCTGTGATAGACAATAACACTAGCAATGAAGAAGCAAAAAATCAA
 AGAAGAAAAATCCAATAAATCCCAAGGAGATTATACGGACTCATTGTGTAATAAAACACAGAAAAATCCCAAAAA
 30 AGAAGATAAAGTTGTCTATATTGCTGAATTTAAAGATAAAGAATCTGGAGAAAAAGCAATCAAGAACTATCCAG
 TCTTAAGAATACAAAAGTTTTATATACTTATGATAGAATTTTAAACGGTAGTGCCATAGAAACAACCTCCAGATAAC
 TTGGACAAAATTAACAAAATAGAAAGTATTTTCATCGGTTGAAAGGGGCACAAAAAGTCCAACCCATGATGAATCAT
 GCCAGAAAAGGAAATTGGAGTTGAGGAAGCTATTGATTACCTAAAGTCTATCAATGCTCCGTTTGGGAAAAATTTT
 GATGGTAGAGGTATGGTCAATTTCAAAATATCGATCTGGAACAGATTATAGACATAAGGCTATGAGAACTCGATGAT
 GATGCCAAAGCCTCAATGAGATTTAAAAAGAAGACTTAAAGGCACTGATAAAAAATTATTGGTTGAGTGATAAAA
 35 ATCCCTCATGCGTTCAATTATTATAATGGTGGCAAAATCACTGTAGAAAAATATGATGATGGAAGGGATTATTTTG
 ACCCAGATGGGATGCATATTGCAGGGATTCTTGCTGGAAATGATACTGAACAAGACATCAAAAACTTAAACGGCA
 TAGATGGAATTGCACCTAATGCACAAATTTCTCTTACAAAATGTATTCTGACGCAGGATCTGGGTTTGGCGGTGA
 TGAACAATGTTTCATGCTATTGAAGATTCTATCAACACACACGTTGATGTTGTTTCGGTATCATCTGGTTTTACA
 GGAACAGGTCCTGTAGGTGAGAAAATATTGGCAAGCTATTGGGCATTAAAGAAAAGCAGGCATTCCAATGGTTGTC
 40 GCTACGGGTAACATATGCGACTTCTGCTTCAAGTCTTCTCATGGGATTAGTAGCAAAATAATCATCTGAAAAATACCG
 AACTGGAAATGTAAACGAACTGCAGCAGATGAAGATGCGATAGCGGTCGTTCTGCTAAAAATCAAAACAGTTG
 AGTTTGATAAAGTTAAACATAGGTGGAGAAAGTTTTAAATACAGAAATATAGGGGCTTTTTCGATAAGAGTAAAA
 TCACAACAAATGAAGATGGAACAAAAGCTCTAGTAAATTAATAATTTGATATATAGGCAAGGGGCAAGACCAAG
 ATTTGATAGGTTGGATCTTAGGGGCAAAATTCAGTAATGGATAGAATTTATACAAAGGATTAAAAAATGCTTT
 45 TAAAAAGCTATGGATAAGGGTGACGGGCCATTATGGTTGTAATACTGTAATTACTACAATAGAGATAATTG
 GACAGAGCTTCCAGCTATGGGATATGAAGCGGATGAAGGTACTAAAAGTCAAGTGTTCATTTCAATTCAGGAGATGA
 TGGTGTAAAGCTATGGAACATGATTAATCCTGATAAAAAAACTGAAGTCAAAAGAAATAATAAGAAAGATTTTAA
 AGATAAATTGGAGCAATACTATCCAATTGATATGGAAGTTTTAATTCCAACAAACCGAATGTAGGTGACGAAAA
 AGAGATTGACTTTAAGTTTGACCTGACACAGACAAAGAACTCTATAAAGAAGATATCATCGTTCCAGCAGGATC
 50 TACATCTTGGGGGCAAGAATAGATTACTTTTAAAAACCCGATGTTTCAGCACCTGGTAAAAATATTAATCCACG
 CTTAATGTTATTAATGGCAATCAACTATGGCTATATGTCAGGAAGTATGAGGCACTCCAATCGTGGCAGCTT
 CTACTGTTTTGATTAGACCGAAATTAAGGAAATGCTTGAAGACCTGTATTGAAAAATCTTAAAGGAGATGACA
 AAATAGATCTTACAAGTCTTCAAAAAATGGCTTCAAAAAATCTGCGCGACCTATGATGGATGCAACTTCTTGGAA
 55 AAGAAAAAGTCAATACTTTGCATCACCTAGACAACAGGGAGCAGGCCTAATTAATGTGGCCAATGCTTTGAGAA
 ATGAAGTTGTAGCACTTCAAAAAACCTGATTCTAAAGGTTTGGTAAACTCATATGTTTCCATTTCTTTAAAGA
 AATAAAAAGGTGATAAAAAATACCTTACAATCAAGCTTCAAAATACATCAAAACAGACCTTTGACTTTTAAAGTTTCA
 GCATCAGCGATAACTACAGATTCTCTAACTGACAGATTAACAACTTGATGAAACATATAAAGATGAAAAATCTCCA
 GATGGTAAGCAAAATTGTTCCAGAAATTCACCCAGAAAAAGTCAAAGGAGCAAAATATCACATTGAGCATGATACT
 60 TTCATATAGGCGCAAAATCTAGCTTTGATTTGAATGCGGTTATAAATGTTGGAGAGGCCAAAAACAAAAATAAA
 TTTGTAGAATCATTATTCTTTGAGTCAGTGAAGCGGATGGAAGCTTAAACTCCAGCGGGAAGAAAAATAAAC
 TTCCAACCTTCTTTGTCGATGCCTCTAATGGGATTGCTGGGAATTGGAACCAACCAATCTTTGATAAATGGG
 CTTGGGAAGAAAGGTTCAAGATCAAAACACTGGGAGGTTATGATGATGATGGTAAACCGAAAAATCCAGGAACCT
 TAAATAAGGGAATTGGTGGAGAACATGGTATAGATAAATTAATCCAGCAGGAGTTATACAAAATAGAAAAAGATA
 65 AAAATACAACATCCCTGGATCAAAATCCAGAATTTGCTTTCAATAACGAAGGATCAACGCTCCATCAATCA
 GTGGTTCTAAGATTGCTAACATTTATCCTTTAGATTCAAAATGGAATCCTCAAGATGCTCAACTTGAAGAGGATT

AACACCTTCTCCACTTGTATTAAGAAGTGCAGAAGAAGGATTGATTTCAATAGTAAATACAAATAAAGAGGGAGA
 AAATCAAAGAGACTTAAAGTCATTTCCGAGAGAACACTTTATTAGAGGAATTTTAAATCTTAAAGCAATGATGC
 AAAGGGAATCAAATCATCTAACTAAAGTTTGGGGTGAAGTGAGGATGGAAGTCACTATAATCTAGAGG
 TAGAGAAGAAAATGCACCAGAAAGTAAGGATAATCAAGATCCTGCTACTAAGATAAGAGGTCAATTTGAACCGAT
 5 TGGCGAAGGTCAATATTTCTATAAAATTTAAATAGATTAACTAAAGATTACCCATGGCAGGTTTCTATATTCCT
 GTAAAAATTGATAACACCGCCCTAAGATTGTTTCCGGTTGATTTTCAAATCCTGAAAAATTAAGTTGATTACAA
 AGGATACTTATCATAAGGTAAAAGATCAGTATAAGAATGAAACGCTATTTCGAGAGATCAAAAAGAACATCCTG
 AAAAATTTGACGAGATTGCGAACGAAGTTTGGTATGCTGGCGCCGCTCTTGTAAATGAAGATGGAGAGGTTGAAA
 10 AAAATCTTGAAGTAACTACGCAGGTGAGGGTCAAGGAAGAAATAGAAAACTTGATAAAGACGGAATACCAATT
 ATGAAATTAAGGTGCGGGAGATTAAAGGGAAAAATCATTGAAGTCATTGCAATTAGATGGTTCTAGCAATTTCA
 CAAAGATTCATAGAAATTAATTTGCTAATCAGGCTGATGAAAAGGGGATGATTTCTATTATCTAGTAGATCCTGA
 TCAAGATTCATCTAAATATCAAAAGCTTGGCGAGATTGCGAATCTAAATTTAAAAATTTAGGAAATGGAAAAAGA
 GGGTAGTCTAAAAAAGATACAACTGGGGTGAACATCATCAAGAAAAATGAAGAGTCTATTAAAGAAAAAT
 15 CTAGTTTTACTATTGATAGAAATTTCAACAATTAGAGACTTTGAAAAATAAGACTTAAAGAAATCATTAAAAA
 GAAATTTAGAGAAGTTGATGATTTTACAAGTGAAGTGGTAAAGAAATGGAGGAATACGATTATAAATACGATGA
 TAAAGGAAATATAATAGCCTACGATGATGGGACTGATCTAGAATATGAACTGAGAACTTGACGAAATCAAATC
 AAAAATTTATGGTGTCTAAGTCCGTCTAAAGATGGCACTTTGAAATCTTGGAAAGATAAGTAATGTTTCTAAA
 AATGCCAAGGTATATTATGGGAATAACTATAAATCTATAGAAATCAAAGCGACCAAGTATGATTTCCACTCAAAA
 ACGATGACATTTGATCTATACGCTAATATTAAATGATATTGTTGGATGGATTAGCTTTTGCAGGAGATATGAGATTAT
 20 TTGTTAAAGATAATGATCAGAAAAAGCTGAAATTAATTAAGATGCCTGAAAAATTAAGGAAATCAAAATCAG
 AATATCCCTATGTATCAAGTTATGGGAATGTCAATAGAAATAGGGGAAGGAGATCTTCAAAAAACAACAGACA
 ATTTAACTAAAAATGGAATCTGGTAAAATCTATTCTGATTGAGAAAAACAACAATATCTGTTAAAGGATAATATCAT
 TCTAAGAAAAAGGCTATGCACTAAAGTGACTACCTATAATCCTGGAAAAACGGATATGTTAGAAGGAAATGGAGT
 CTATAGCAAGGAAGATATAGCAAAAAATACAAAAGGCCAATCTAATCTAAGAGCCCTTTCAGAAACAACAATTTA
 25 TGCTGATAGTAGAAATGTTGAAGATGGAAGAAGTACCCAATCTGTATTAATGTCCGCTTTGGACGGCTTTAAACATT
 ATAAGGTATCAAGTGTTTACATTTAAATGAACGATAAAGGGGAAGCTATCGATAAAGACGGAATCTTGTGACA
 GATTCCTCTAACTTGTATTATTTGGTAAGGATGATAAAGAATACACTGGAGAGGATAAGTTCAATGTGAAGAGCTA
 TAAAGAGAAGATGGCTCCATGTTATTTATTGATACCAACCAAGTAAACCTTTCAATGGATAAGAACTACTTTAATCC
 ATCTAAATCTAATAAAATTTATGTACGAAATCCAGAATTTTATTTAAGAGGTAAGATTTCTGATAAGGGTGGTTTT
 30 AACTGGGAATTGAGAGTTAATGAATCGGTTGTAGATAATTATTTAATCTACGGAGATTACACATTGATAACACTA
 GAGATTTAATATTAAAGCTGAATGTTAAAGACGGTGACATCATGGACTGGGAATGAAAGACTATAAAGCAAAACG
 GATTTCCAGATAAGGTAACAGATATGGATGGAATGTTTATCTTCAAACTGGCTATAGCGATTGTAATGCTAAAGC
 AGTTGGAGTCCACTATCAGTTTTATATGATAATGTTAAACCCGAAGTAAACATTGATCCTAAGGGAAATACTAGT
 ATCGAATATGCTGATGGAATCTGTAGTCTTAAACATCAATGATAAAGAAATAATGGATTGATGGTGAGATT
 35 CAAGAACAACATATTTATATAAATGGAAAAAGATATACATCATTTAATGATATTAAACAAATAATAGACAGACA
 CTAACATTAAGATTGTTGTAAGATTTTGAAGAAATACAAACCGTAAAAGAATTCAATTTAAATAAAGATACG
 GGAGAGGTAAGTGAATTAACCTCATAGGGAATCTGTGACCATTCAAATGGAAGAAATGAGTTCAACGATA
 GTGTCCGAAGAAGATTTTATTTTACCTGTTTAAAGGTGAATTAAGAAAAAGGATACCAATTTGATGGTTGGGAAA
 40 TTTCTGGTTTTCGAAGGTAAAAAGACGCTGGCTATGTTATTAATCTATCAAAAGATACCTTTATAAAACCTGATT
 CAAGAAAAATAGAGGAGAAAAAGGAGGAAGAAAAATAAACCTACTTTTATGATGATCGAAAAAGAAAGATAACCCAC
 AAGTAAACCATAGTCAATTAATGAAAGTACAGAAAAAGAGGATTTACAAAGAGAAGAGCATTACAAAAATCT
 GATTCAACTAAGGATGTTACAGCTACAGTCTTGATATAAAAAACAATATCAGTAGTAAATCAACTAAACATCT
 AATAAGTTGCCAAAACTGGAACAGCAAGCGGAGCCAGACACTATTAGCTGCCGAATAATGTTTATAGTAGGA
 45 ATTTTTCTGGATTGAAGAAAAAAATCAAGATTAA

 YPVVLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVSKNPVIDNNTSNEEAKIKEENS
 KSQGDYDTSFVNKNTENPKKEDKVYVIAEFKDKESGEKAJELSSLKNTKVLTYDYRIFNGSAIETTPDNLKIKQIEG
 SVERAQKVQPMNHNHARKEIGVEEADYLSINAPFGKNFDRGMVISNIDTGTDRHKAMRIDDDAKASMRFKKEDL
 50 KGTDKNYWLSDKIPAFNYNNGGKITVEKYDDGRDYFDPHGMHIAIGLAGNDEQDIKNFNGIDGLAPNAQIFSYKMY
 SDAGSGFAGDETMFHAIEDSIKHNVVSVSSGFTGTGLVGEKYWQAIRLRKAGIPMVVATGNVYATSASSSSWDLVA
 NNHLKMTDTGNVTRTAHEDAIASAKNQTFEDKVNIGGESFKYRNIGAFFDKSKITTNEDGTAKPSKLFVYIGK
 GQDQDLIGLDLRGKIAVMDRIYTKDLKNAFFKAMDKGARAJMVNTVNYNVRDNTWELPAMGYEADGTSQVFSI
 SGDDGVKLVNMINPDKKTEVKRNKEDFKDKLEQYYPIDMESFNSNKPNGDEKIDFKFAPDTEKELYKEDIIVPAG
 55 STSWGPRIDLLKPDVSAPEGKNIKSTLNVINGKSTYGYMSGTSMATPIVAASTVLIRPKLEMLERPVLKLNKGGDDKIDL
 TSLTKIALQNTARPMMDATSWKEKSQYFASPRQAGLINVANALRNEVVATFKNTDSKGLVNSYGSISLKEIKGDKK
 YFTIKLHNTSNRPLTFKVSASAITDSDLRLKLDDEYKDEKSPDGKQIVPEIHPEKVKGANITFEHDTFTIGANSSFDLN
 AVINVGEAKNKNKFVESFIHFESVEAMEALNSSGKKINFQPSLSMPLMGFAGNWNHEPILDKWAWEEGSRSKTLGGYD
 DDGPKPIPGTLNKGIGGEHGDKFNPAQVIQNRKDKNTSLDQNPFLFAFNNEGINAPSSSGSKIAMYPLDSNGNPQDA
 60 QLERGLTPSPLVLRSAEEGLISIVNTNKEGENQRDLKVISREHFIRGILNSKSNDAKGKSSKLKVWGDLLKWDGLYNPRG
 REENAPESKDNQDPATKIRGQFEPIAEGQYFYKFKYRLTKDYWPQVSYIPVKIDNTAPKIVSVDFSNPEKIKLITKDTYHK
 VKDQYKNETLFARDQKEHPEKFDEIANEVWYAGAALVNEDEGEVEKNLEVTYAGEGQGRNRKLDKDGNTIYEIKGAG
 DLRGKIEVIALDGSNFTKIHRIKFANQADEKGMISYYLVDPPDQDSSKYQKLGEIAESKFKNLNGKEGSLKKDDTTGVE
 HHHQENEEISIEKSSFTIDRNISTIRDENKDLKKLKKKREVDFTSETGRMEEYDYKYDDKGNIIAYDDGTDLEYE
 65 TEKLDEIKSKIYGLVSPSKDGHFEILGKISNVSKNAKYVYGNKYKIEIKATKYDFHSKTMTFDLYANINDIVDGLAFAG
 DMRLFVKDNDQKKAIEKIRMPKIKETKSEYPYVSSYGNVIELGEGDLSKNKPDNLTKMESCKIYSDSEKQYLLKDNII

5 L R K G Y A L K V T T Y N P G K T D M L E G N V Y S K E D I A K I Q K A N P N L R A L S E T T I Y A D S R N V E D G R S T Q S V L M S A L D G F N H R Y Q
V F T F K M N D K G E A I D K D G N L V T D S S K L V L F G K D D K E Y T G E D K F N V E A I K E D G S M L F I D T K P V N L S M D K N Y F N P S K S N K I
Y V R N P E F Y L R G K I S D K G G F N W E L R V N E S V D N Y L I Y G D L H I D N T R D F N I K L N V K D G D I M D W G M K D Y K A N G F P D K V T D
M D G N V Y L Q T G Y S D L N A K A V G V H Y Q F L Y D N V K P E V N I D P K G N T S I E Y A D G K S V F N I N D K R N N G F D G E I Q E H I Y I N G K
E Y T S F N D I K Q I I D K T L N I K I V V K D F A R N T T V K E F I L N K D T G E V S E L K P H R V T V T I Q N G K E M S S T I V S E E D F I L P V Y K G E L E K
G Y Q F D G W E I S G F E G K K D A G Y V I N L S K D T F I K P V F K I E E K E E E N K P T F D V S K K C D N P Q V N H S Q L N E S H R K E D L Q R E E H
S Q S D S T K D V T A T V L D K N N I S S K S T T N P N K L P K T G T A S G A Q T L L A A G I M F I V G I F L G L K K N Q D Z

10 ID6 597 bp

10 C T T G A A T T A A A T A A A A A C G T C A T G C G A C T A A G C A T T T T A C T G A T A A G C T T G T T G A T C C A A A G A T G T G C G T A C G G
C T A T C G A A A T T G C A A C C T T A G C G C C A A G C G C C A C A A C A G C C A G C C T T G G A A A T T T G T G G T G G T A C G T G A G A A A A
A T G C T G A A C T G G C A A A G T T A G C T T A T G G T T C C A A T T T T G A A C A G G T A T C A T C A G C G C C T G T A A C C A T T G C C T T G T T
T A C A G A T A C G G A C T T A G C C A A C G T G C T C G T A A G A T T G C C C G T T G T T G G T G G T A A T A A C T T T T C T G A A G A G C A A
C T T C A A T A T T T T A T G A A A A T C T G C C A G C T G A G T T T G C C C G T T A C A G T G A G C A A C A A G T C A G C G A C T A C C T A G C T C
T C A A T G C A G G T T T G G T T G C C A T G A A C T T G G T T G C A T T G A C A G A C C A A G G A A T T G G T T C T A A C A T T A T T C T T G G
T T T T G A C A A A T C A A A A G T T A A T G A A G T T T T G A A A T C G A A G A C C G T T T C C G C C C A G A A C T T G A T C A C A G T G G G T
T A T A C A G A C G A A A A T T G G A A C C A A G C T A C C G C T T G C C A G T A G A T G A A A T C A T C G A G A A A A G A T A G

20 L E L N K K R H A T K H F T D K L V D P K D V R T A I E I A T L A P S A H N S Q P W K F V V V R E K N A E L A K L A Y G S N F E Q V S S A P V T I A L F T D T
D L A K R A R K I A R V G G A N N F S E E Q L Q Y F M K N L P A E F A R Y S E Q Q V S D Y L A L N A G L V A M N L V L A L T D Q G I G S N I L G F D K S K
V N E V L E I E D R F R P E L L I T V G Y T D E K L E P S Y R L P V D E I I E K R Z

25 ID7 1401 bp

25 A T G A C A G C A A T T G A T T T T A C A G C A G A A G T A G A A A A C G C A A A G A A G A C C C T C T T G G C T G A C T T G T T A G C C T T T T G
G A A A T C A A T T C A G A A C G T G A T G A C A G C A A G G C T G A T G C C C A G C A T C C A T T T G G G C C T G G T C C A G T A A A A G C C T T G
G A G A A A T T C C T T G A A A T C G C A G A C C G C G A T G G C T A C C C A A C T A A G A A T G T T G A T A A C T A T G C A G G A C A T T T T G A G
T T T G G T G A T G G A G A A G A G T T C T C G G A A T C T T T G C C C A T A T G G A T G T G G T G C C T G C T G G T A G C G G T T G G G A C A C A G
30 A C C C T T A C A C A C C A A C T A T C A A A G A T G G T C G C C T T T A T G C G C G C G G G C T T C G G A C G A T A A G G G T C C T A C A A C A G
C T T G T T A C T A T G G T T T G A A A A T C A T C A A A G A A T T G G G T C T T C C A A C T T C T A A G A A A G T T C G C T T C A T C G T T G G A A C
A G A C G A A G A A T C A G G C T G G G C A G A C A T G G A C T A C T T T G A G C A C G T A G G A C T T G C C A A C C A G A T T T C G G T T T
C T C A C C A G A T G C T G A A T T T C C A A T C A T C A A T G G T G A A A A G G A A A T A T C A C G G A A T A C C T C C A C T T T G C A G G A G A
A A A T A C A G G T G T T G C C C G T C T T C A C A G C T T T A C A G G T G G T T A C G T G A A A A T A T G G T A C C A G A A T C A G C A A C A G C
35 A G T C G T T C A G G T G A C T T G G C T G A C T T G C A A G C T A A A C T A G A T G C C T T T G T T G C A G A A C A C A A C T T A G A G G A G A
A C T C C A A G A A G A A G C T G G C A A A T A C A A G G T G A C G A T C A T T G G T A A A T C A G C C C A C G G T G C T A T G C C T G C T T C A G G
T G T C A A T G G C G C A A C T T A C C T T G C C C T C T C C T C A G C C A G T T T G G C T T T G C T G G T C C A G C C A A G A C T A C C T T G A C
A T C G C A G G T A A A A T T C T T T G A A C G A T C A T G A G G G T G A A A A T C T T A A G A T T G C T A T G T G A T G A A A A G A T G G G T
G C T C T T T C T A T G A A T G C C G G C G T C T T C C A C T T C G A T G A A A C A A G T G C T A A T A C C A T T G C C C T A A C A T C G G C T
40 A T C C A A A G G A A C A A G T C C A G A A C A A T C A A G T C A A T C C T T G A A A A C T T G C C A G T T G T T T C T G T A G C C T G T C T G A
A C A C G G T C A C A C G C C T C A C T A T G T G C C A A T G A A G A T C C A C T T G T G C A A C C T T G T T G A A T A T C T A T G A A A A C A
A A C T G G C T T T A A A G G T C A T G A A C A A G T C A T C G G T G G T G G A A C C T T T G G T C G C T T G C T A G A A C G C G G A G T T G C C T A
C G G T G C T A T G T T C C C A G A C T C G A T T G A T A C C A T G C A C C A A G C C A A T G A A T T A T C G C C T T G G A T G A T C T T T C C G A
45 C G A C G A C C A A T T A T G C C G A A G C T A T T T A C G A A T T G A T C A A A T A A

50 M T A I D F T A E V E K R K E D L L A D L F S L L E I N S E R D D S K A D A Q H P F G P G P V K A L E K F L E I A D R D G Y P T K N V D N Y A G H F E G D
G E E V L G I F A H M D V V P A G S G W D T P Y T P T I K D G R L Y A R G A S D D K G P T T A C Y Y G L K I I K E L G L P T S K K V R F I V G T D E E S G W
A D M D Y Y F E H V G L A K P D F G F S P D A E F P I I N G E K G N I T E Y L H F A G E N T G V A R L H S F T G G L R E N M V P E S A T A V V S G D L A D L
Q A K L D A F V A E H K L R G E L Q E E A G K Y K V T I I G K S A H G A M P A S G V N G A T Y L A L F L S Q F G F A G P A K D Y L D I A G K I L L N D H E G
E N L K I A H V D E K M G A L S M N A G V F H F D E T S A D N T I A L N I R Y P K G T S P E Q I K S I L E N L P V V S V L S E H G H T P H Y V P M E D P L V Q
T L L N I Y E K Q T G F K G H E Q V I G G G T F G R L L E R G V A Y G A M F P D S I D T M H Q A N E F I A L D D L F R A A A I Y A E A I Y E L I K Z

ID8 1617 bp

55 G T G T A T A C T A T T A T A A A A T C A A A T A T A A A A A A A T T A G T T T A T T A A C G A T A T T T A T T G T T G C T G G T C A A T T A T T G C T
A A T T T A T G C A G C A A C T A T T A A T G C T C T G G T G T T G A A T G A A T T A A T T G C G A T G A A T T A G A G C G G T T T T G A A A T T G
T C A A T C T A C C A A A T G A T T G T C T G G T G T G G G A T A A T T C C T T G A C T G G G T A G T G A A A A A T T A C A G G T T G A A G T G A
T C C A A G A G T T T A A T C T A G A G A T T C G A A A T A G A G T T G C C A C A G A C A T C T A A C T C T A C C T A T C A A G A A T T T C A T A G
60 T A A A T C A T C A G G A A C A T A T C T T C G T G G C T A A A T A A T G A T G T T C A G A C T T T A A A T G A T C A G G C G T T T A A A C A C T T
T T T T T A G T A A T A A A A G G A A T T T C T G G T A C T A T A T T T G C A G T T G T G A C T C T T A A T C A C T A T C A T T G G T C A T T G A C T G T
A G C C A C C T T G T T T C A T T A A T G A T T A T G C T A C T G T A C C A A A A T C T T T G C A T C G A A A A T G C G A A G T T A G T C T A
A A T T T A A C T A A C C A A A A T G A A G C T T T T T A A A A T C T A G T G A G A C T A T A T T G A A T G G A T T T G A T G T T A G C G C C T
T G A A T C T T T A T A T G T A T T G C C T A A G A A A A T T A A A G A A G C A G A A T T T A T T A A A G A T G G T T A T A C A A G A A A G A
C A A C T G T A G A A A C G T T A G C A G G C G C T A T A G C T T C T T C T C A A T A T T T T T T C A G A T A T C T C T G T T T T T T A A C A
65 G G C T A T C T G C A A T A A A A G G A A T A G T G A A A A T T G G T A C T A T T G A A G C A A T A G G A G C A C T A A C A G G T G T T A T T T T

5 ACAGCGCTAGGTGAATTAGGAGGTCAATTATCCTCTATTATTGGTACGAAGCCTATTTTTTAAAAATTGTATTCAA
TTAATCCAATTGAGTCAAATAAAATGAATGATATCGAACCAATGAGGTGAATAGAGATTTCCGTTATATGAAG
CAAAAAATATTTGCTATAAGTATGGAGATAAAGAAAATTTAAAAAACTTAAATTTTTGTTTCAACGTAATGAAAA
GTATTTAATTTTAGGTGAAAGTGAAGCGGGAATCTACATTATTAATAATTGAATGGCTTTTGGAGAGATTAT
10 AGTGGAGAATTGCGATTCTGCGGGGATGATATAAAAAAACTCCTATTATAATATGGTTTCGAATGTTCTATATG
TAGATCAAAAAGCTTATTGTTTGAAGGTACGATTAGAGATAATTTTTATTGGAAGAAAAATTACTGATGAAGA
AATACTACAGTCTTTAGAGCAAGTTGGTTTGAGTGTAAGATTTTCTAATAACATTTTAGATTATTATGTTGGT
GATGATGGGAGATTACTGTGAGGAGGCGAGAAACAAAAAATTACTTAGCTAGAGGGCTAATTAGAAAAAGAA
AATAGTATTAATTGACGAGGGAACCTCTGCTATCGATAGGAGAACTTCGTTAGCGATTGAACGTAAGATATTAGA
TAGAGAGGATTTGACTGTCAATTATTGTTACCCATGCTCCGCATCCGGAACCTTAAACAATATTTTACTAAGATATAT
CAATTTCCAAAGGATTTTATTTAA

15 MYTIKSNKKFSLLTIFIVAGQLLLIYAATINALVLNELIAMNLERFLKLSIYQMIVWCGIIFLDWVVKNYQVEVIQEFNL
EIRNRVATDISNSTYQEFHKSSTGYLSWLNNDVQTLNDQAFKQLFLVKGISGTFIVVTLNHYHWSLTVATLFSLMIM
LLVPKIFASKMREVSLLNTQNEAFLKSETILNGFDVLASLNLVYLPKKIKEAGILLKMMVIQRKTTVETLAGAISFFLNI
FFQISLVFLTGYLAIKGIVKIGTIEAIGALTGVFTALGELGGQLSSIGTKPIFLKLYSINPIESNKMNDIEPNEVNRDPLYE
AKNICYKYGDKEILKNLNFQFORNEKYLILGESGSGKSTLLKLLNGFLRDYSGELRFGDDIKKTSYLNMVSNVLVYDQ
KAYLFEGTIRDNILLEENYTDDEILQSLEQVGLSVKDFPNNILDYVVGDDGRLLSGGQKQKITLARGLRNKKIVLIDEGT
20 SAIDRRTSLAIERKILDREDLTVIIVTHAPHELKQYFTKIYQPKDFIZ

ID9 705 bp

25 ATAACAGTTAAACAGATTATGGACGAAATAGCCGTTTCAGATATGACTGCAAGGCGCTATTTACAGGAATTAGCT
GATAAAGATTGCTGATTGCTGTGTCATGGTGGAGCTGAAAAAACTTCGAACCAACTCCCTTTTGCTAATGAGCGAT
CAAATATTGAAAAACAAGCCCTCCAAACGGCAGAAAAACAAGAAATAGCCCATTTTGCAGGCAGTCTAGTAGAA
GAAAGAGAACTATTTTCATTGGACCGGAAACAACATTAGAGTTTTTTCGCGGTGAGTTGCCTATTGACAATATCC
CGCTCGTAACCAACAGTCTACCTGTTTTCTGATTTAAGCGAAGCAAAAAATTAACAGATTGATTTTAAATAGGTGG
AAATTATCGCGATATTACAGGTGCTTTTGTGGTACATTGACCTTACAAAAATCTCTAATCTCCAATTTTCTAA
30 GCTTTCTGTTAGCTGTAATGGTATTCAAAACGGAGCTCTAGCTACTTTAGCGAGGAAGAGGGAGAGGCTCAACGC
ATCGCTTTAAATAATTCTAATAAAAAATATTTACTCGCAGATCATAGCAAGTTCAATAAGTTTGATTTTTATACTTT
TTATAATGTATCAAACTTGATACTATTGTTTCAGATTCTAAACTAAGTGATTCAATCCTTTTTAAGCTATCTAAAC
ACATTAAGTCATCAAGCCTTAA

35 ITVKQIMDEIAVSDMTARRYLQELADKDLLIRVHGAELRNTNSLLTNSRNIQALQTAEKQEIHFAGSLVEERETI
FIGPQTLEFFARELPIDNIRVVTNSLPVFLILSERKLTDLILIGGNYRDITGAFVGTLLQNLNLQFSKAFVSCNGIQNGA
LATFSEEEGEAQRIALNNSNKKYLLADHSKFNFDFYTFYNVSNLDTTVSDSKLSDSILFKLSKHVKIPZ

ID10 483 bp

40 ATGACTGAGTTTTGTTAGATCTTCTCTAGAACCCATTAACTAGCTCGTTGGACCTACTACTATCACTTGAAAC
AGCTAGACAAAAACAGATAAAGACCAAGAGCTTAAACTGAAATTCATCCATCTTTATCGAACACAAGGGAAATT
ATGCTTATCGCCGGGTTCATTTAGAACTAAGAAATCGTGTTATCTGGTAAATCATAAAAAGAGTTCAAGGCTTGAT
GAAAGTACTCAATTTACAAGCTAAATGCGAAAGAAACGAAATATTCTTCTCATAAAGGAGACGTTGGTAAAGAA
45 GGCAGAGAATCTCATTCAAGCCCAATTTGAAGGCTCTAAAACAATGGAAGAGTGCTACACAGATGTGACTGAATT
TGCCATTCCAGCAAGTACTCAAAAGCTTTAGCTATCACCAGTTTATGATGGCTTTAACAGCGAAATTATTGCTTTT
AATCTTCTTGTTCGCTAATTTAGAATAA

50 MTEFSLDLLLEAIKLARWYTYHLLKQLDKTDKQDELKTEIQSIFIEHKGNYA YRRVHLELRNRGYLVNHKRVQGLMK
VLNLQAKMRKKRYSSHKGDVGKKAENLIQAQFEGSKTMEKCYTDVTEFAIPASTQKLYLSPVLDGFNSEIAFNLS
PNLEZ

ID14 1266 bp

55 CCAGGATTTGGTACCGTTGCAAGTGGTGTGCCTTTCTCTCTAAAGGAAAAATGGAGGAAAAATCAATCAATCAGCA
CATTCAGATATCAAAGTTGCTAAGGTATTGGTCAAGGATGAAGATGAAAAAATCGCTTGCTTGCAAGGGAAT
GACTTTAACTTTGTAACCAATGTGGATGATATTTATCAGACCAAGGATATTACTATCGTAGTGCAATTGAGGGC
GTATTGAGCCTGTCTAAAACCTTTATCACTCGTGCCTTGGAAAGCTGGAAACACGTTGTTACTGCTAAACAGGACCT
TTAGCTGTCCATGGCGCAGAATTGCTAGAAATCGCTCAAGCTAACAAGGTAGCACTTTACTACGAAGCAGCAGT
60 TGCTGGTGGGATTCCAATTCTCGTACTTTAGCAAAATTCCTTGGCTTCTGATAAAATTACGCGCGTGTCTGGAGTA
GTCAACCGAACTTCCAATTCATGGTGACCAAGATGGTGAAGAAGGCTGGTCTTACGATGATGCTCTTGGCGAA
GCACAACGTCTAGGATTTGCAGAAAGCGATCCGACGAATGACGTAGATGGGATTGATGCAGCTACAAGATGGTT
ATTTTGAAGCAATTTGCTTTGGCATGAAGATTGCTTTGATGATGTAGCCACAAGGGAATCCGCAATATCACAC
CAGAAGACGTAGCTGTAGCTCAAGAGCTTGTTACGTAGTGAAATTGGTTGGTTCTATTGAGGAACTTCTTCAGG
TATTGCTGCAGAGTGACTCCAACCTTCTACCTAAAGCGCACCCACTTGCTAGTGTGAATGGCGTAATGAACGCT
65 GTCTTTGTAGAATCTATCGGTATTGGTGAGTCTATGTAACGACCAAGGTGCGGGTCAAAAACCAACTGCAACA

5 AGTGTGTAGCTGATATTGTCCGTATCGTTCGTGTTGAATGATGGTACTATTGGCAAAGACTTCAACGAATATA
GCCGTGACTTGGTCTTGGCAAATCCTGAAGATGTCAAAGCAAACCTACTATTCTCAATCTTGGCTCTAGACTCAA
AGGTCAGGTCTTGAAGTTGGCTGAAATCTTCAATGCTCAAGATATTTCTTTAAGCAAATCCTTCAAGATGGCAA
GAGGGTGACAAGGCGCGTGTCTTATCATCACACAAGATTAAATAAGCCAGCTTGAAAAATGTCTCAGCTGAA
TTGAAGAAGGTTTCAGAATTCGACCTCTTGAATACCTTCAAGGTGCTAGGAGAATAA

10 PGFGTVASGVPFLKENGKINQSAHSDIKVAKVLKDEDEKNRLLAAGNDFNFVTNVDDILSDQDITTVELMGRIEP
AKTFITRALEAGKHVVTANKDLLAVHGAELLEIAQANKVALYYEAAVAGGIPILRTLANSLASDKITRVLGVVNGTSNF
MVTKMVEEGWSYDDALAEQRLGFAESDPTNDVDGIDAAKYMVLSQFAFGMKIAFDDVAHKGIRNITPEDVAVAQE
LGYYVKLVGSIEETSSGIAAEVPTFLPKAHLASVNGVMNAVFVESIGIGESMYGPGAGQKPTATSVVADIVRIVRRL
NDGTIGKDFNEYSRDLVLNPNEDVKANYYSILALDSKGQVLKLAEIFNAQDISFKQLQDQKEGDKARVVITHKINKA
QLENVSAELKKVSEFDLLNTFKVLGEZ

15 ID16 1725 bp

ATGAAACACCTATTATCTTACTTCAAACCTACATCAAGGAATCAATTTTAGCCCCCTTGTTCAGCTGTTAGAAG
CTGTTTTGAGCTCTTGGTTCCTCATGGTGAATGCTGGGATTGTTGACCAATCTTTACCTCAGGAGAGTCAAGGTCA
TCTCTGGATGCAGATTGGCCTGCTCCTTATCTTGCAGTAATGGCGTCTTGTAGTGGCCTTGATAGCTCAATTTTACT
CAGCAAAGGCAGCAGTAGGTTCTGCTAAGGAATTGACAAACGATCTTTATCGTCATATTTCTTCTGCCCCAAGGA
CAGCAGAGACCGTCTGACAACTTCTAGTTTGGTCACTCGCTTGAATTCGGATACCTACCAGATTCAAGCTGGTATC
AATCAATTCCTGCGTCTCTTTTACGAGCGCCATTATCGTTTTGGTGCCATTTTATGGCTTATCGAATCTCAGC
TGAGTTGACTTTCTGGTCTTATGCTTGGTTCATTTGACCATTTGTCATTGTAGGGTTATCTCGATTGGTCAATC
CTTTCTACAGTAGTCTCAGAAAGAAACGGACCAACTGGTTCAGGAAACGCGCCAGCAATTGCAAGGGATGCGGG
TTATTCTGCTCTTTGGTCAAGAAAAACGAGAGTTACAGATTTTCAAACCTTAACCAAGTTTATGCTAGATTACA
AGAAAAGACAGGTTTCTGGTCTAGTTTATTAACACCTCTGACCTATCTGATTGTCAATGGAACTCTTCTCGTTATT
ATCTGGCAAGGCTATATTTCAATTCAAGGAGGAGTGCTCAGTCAAGGTGCTCTCATTGCTCTTATCAATTACCTCT
TACAGATTTTGGTGGAATTGGTCAAGCTAGCCATGTTGATCAATTCCTCAACCAAGTCCCTATATCTCAGTCAAGCG
AATCGAGGAAGTCTTTGTTGAGGCTCCAGAGGATATCCATTAGAGTTAGAACAAAAAGCAAGCTACCCAGAGATAA
GGTTTTACAAGTCCAAGAAATTGACCTTTACCTATCCTGATGCGGCCCGCCAGCCTTCTCTGAGATACATTTCTTTGAT
ATGACTCAAGGACAAATTCTAGGTATCATCGGGGAACTGTTCTGGTAAATCAAGCTTGGTGCAACTCTTACTTG
GACTTTATCCAGTAGACAAGGGGAACATTGACCTTTATCAAAATGGACGTAAGTCTCTTAATTTGGAGCAGTGGC
GGTCTTGGATTGCTATGTACCTCAAAAGGTGCAACTCTTTAAAGGAACCACTTCTGTTCCAACTGACTTAGGTTT
CAATCAAGAAGTATCTGACCAGGAACCTCTGGCAGGCTTGGAGATTGCGCAAGCTAAGGATTTGTCAAGTAAAA
GGAAAGGACTCTTGGATGCTCTAGTTGAGGCAAGGGGGCGAAATTTCTCAGGTGGACAAAAACAAGATTGTCTAT
CGCCCGAGCAGTCTTGGCCAGGCTCCGTTTCTCATCTAGATGATGCAACCTCGGCACTGGATACCATTTACAGAG
TCCAAGCTCTTGAAGCTATTAGAGAAAAATTTTCCAAACACGAGCTTAATTTTGATCTCTCAACGAACTCAACTT
TACAGATGGCGGACAGATTCTCCTCTTGGAAAAAGGTGAGTTGCTAGCTGTTGGCAAGCAGATGACTTGATGA
AATCCAGCCAAGTCTATTGTGAAATCAATGCATCCCAATGGAAGGAGGACTAG

40 MKHLLSYFKPYIKESILAPFLKLEAVFELLVPMVIAGIVDQSLPQGDQGHLMQIGLLIFAVIGVLVALIAQFYSKA
AVGSAKELTNDLYRHLSLPKDSRDLTSSLYTRLTSDTYQIQTGINQFLRLFLRAPIVFGAIFMAYRISAELTFWFLVL
VAILTIVIVGLSRLVNPFFYSRLKKTDLVQETRQQLQGMVRIRAFQGEKRELIQFOTLNQVYARLQEKTFWSSLLTPL
TYLIVNGTLLVIIWQGYISIQGGVLSQGALLALINYLQILVELVKLAMLINSLNQSYISVKRIEVEVFVEAPEDIHSELEQKQ
45 ATRDKVLQVQELTFTYPDAAQPSLRYSFDMTQGGILGIGGTGSGKSSLVQLLGLYPVDKGNIDLYQNGRSPNLQEQ
WRSWIAVYPQKVELFKGTIRSNLTGTFNQEVSDQELWQALEIAQAKDFVSEKEGLLDALVEAGGRNFGGQKQRLSLIA
RAVLRQAPFLILDDATSLDTITESKLLKAIENFPNTSLILISQRTSTLQMAHQILLLEKGELLA VGKHDLMKSSQVYC
EINASQHGKEDZ

50 ID18 1224 bp

ATGAAACGTTCTCTCGACTCAAGAGTCGATTACAGTTTGCTCTTGCCAGTATTTTTCTACTGGTCATCGGTGTTGGT
GGCTATCTATATAGCCGTTAGTCATGATTATCCCAATAATATTCTGCCATTTTAGGGCAGCAGGTGCGCTGGATT
GCCTTGGGGCTTGTGATTGGTTTGTGGTCATGCTCTTTAATACAGAATTTCTTTGGAAGGTGACCCCTTTCTATA
55 TATTTTAGGCTTGGGACTTATGATCTTGCCGATTGTATTTATAATCCAAGCTTAGTTGCATCAACGGGTGCCAAA
AACTGGGTATCAATAAATGGAATTACCCTATTCCAACCGTCAGAAATTTATGAAGATATCCTATATCCTCATGTTGG
CTCGTGTCTATTGTCATTTACAAAGAAACATAAGGAATGGAGACGCACGGTTCGGCTGGACTTTTGTAAATTTT
CTGGATGATTCTCTTACCATTCCAGTCTAGTTCTTTAGCACTTCAAAGTGACTTGGGGACGGCTTGGTTTTTG
TAGCCATTTCTCAGGAATCGTTTTATTATCAGGGGTTTCTTGGAAATTTATATCCAGTATTTGTGACTGCTGTA
ACAGGAGTTGCTGGTTTCTAGCTATCTTTATTAGCAAGGACGGACGAGCTTTTCTTACCAGATTGGAATGCCGA
60 CTTACCAAAATTAATCGGATTTTGGCTTGGCTCAATCCCTTTGAGTTTGGCCAAACAACGACTTACCAGCAGGCTCA
AGGGCAGATTGCCATTGGGAGTGGTGGCTTATTTGGTCAGGGATTTAATGCTTCAATCTGCTTATCCAGTTTGA
GAGTCAGATATGATTTTACGGTTATTGCGAAGATTTTGGCTTTATTGGCTCTGCTGCTGTTATTGCCCTCTATCT
CATGTTGATTTACCGTATGTTGAAGATTACTCTTAAATCAATAACCAAGTTCTACACTTATATTTCCACAGGTTGA
65 TTATGATGTTGCTCTTCCACATCTTTGAGAATATCGGTGCTGTGACTGGACTACTTCTTTGACGGGGATTCCCTTG
CCTTTCATTTGCAAGGGGGATCAGCTATTATCAGTAATCTGATTGGTGTGGTTTTGCTTTTATCGATGAGTTACCA

MKRSLDSRDVYSLLLPVFFLLVIGVVAIYIAVSHDYPNNILPILGQQVAWIALGLVIGFVVMLENTEFLWKVTPFLYILGL
 GLMILPIVFYNPSLVASTGKANWVSINGITLFQPESEFMKISYILMLARVIVQFTKKHKEWRRTVPLDFLLIFWMILFTTPVL
 VLLALQDGLTALVFVAIFSGVILSGVSWKIIIPVFVTAVTGVAFLAIFISKDGRAFLHQIGMPPTYQINRILAWLNPFEE
 AQTTTYQQAQGGIAGSGGLFGQGFNASNLIPRVESDMIFTVIGLDFGSGVSLVIALYLMLYRMLKITLKSNNQFITY
 ISTGLIMMLLFHIFENIGAVTGILLPTGIPLPFISQGGSAJNSNLIAVGDLSSYQTNLAEEKSGKVPFKRRKVVYLKOKZ

ATGTTGGCTAAGAAAAAAATCTTATTTTTATGGTCTTTTTCTCTTGGAGGTGGTGCAGAGAAGATTCTATCAA
CCATTGTTTCAAATCTGGATCGCAGAAAAGATGATTGATATTCTGAAATGGAGCACTTTGACAAGGGGATATGA
ATCTGTTCCAAGCATGTACGCATTTAAAAATCCCTTCAAGATTATCGCAAAACCGATGGTTACGAGCTTTTTG
TGGAGAATGAGAATTTATTTTCCAAGACTGACTCGTCGTTTCTGTGAAAAGATGATTATGATGTGTGAAGTTTCTT
TTACCATTAATGAATCCACCACCTGTGTTCTCTAAAAAGAGAAGTCAAGAAAGATCTCTGGATTCTGGAATCATGGAAGTAT
TGAAAGAACTCTTAAGGATAGCTCTAAAGAGAAATCATAGAAAGCCAGTTGGATGCTGCGAATCAAAATGTGAGG
GATTTCAAAAAAGACCAGCAATTCTATCAAGGAAGTTTATCCAGATTATACTTCTAAATACAGACAATCTACAAT
GGATATGATTTTCAGACTATTTAGAAAATCTCAAGAGAAGATCGAATATCGAGATGTCTCTCTCAAAGTATCTGTA
CTATCGGACGGATGAGGAAATTAAGGGTTCTGACCGGTGATGGGAAGTGAATTCAGTATTACCAAGAGGGGAA
AAAACTACTATCTCTATTTTATCGGGGCTGGTGATATGGAAGAGGGAAGTGA AAAAACCGAGTCAAAAGATATGGGA
TTGAGGACTATGTACATTTCTCTGGTGTATCAAAAAATCTTATCAGATCTATCTCAGACGAAGGATCTTTTGTCT
ATGTCTAAACAAAGAGGTTTTCTCTGGAGTGTATGGTGAAGCCTTGAAGTCTGGGACTCCCTTTATCTCTACGGACG
TTGAGGGGGCTGAGGAATATCCCAAGAAGGACGATTGGACAAATCATTTGAGAGCAATCAAGAGGCGAGCTCAG
CGCATTTACTAATGACTGACTTGCCTCAAAATCTTGATGTGATAGGGCTAGGCCAATTCATTCAACAATTTACAA
TTACAAAACAAATCGAACAAAGTGA AAAAACTATTAGGAGGATAG

ID23 1434 bp

METALISVIVPVYNVAQYLEKSIASIQKQTYQNLEILVDDGATDESGRLCDSLAEQDDRVSVLHKKNEGLSQARNDGM
KQAHG DY LIFIDSDDYTHPEMIQSLYEQLVQEDADVSSCGVMNVYANDESPQSANQDDYFVCDSTFLKEYLIGEKIPG
TICNKLKRQIATALSFPKGLIYEDAYYHFDLKLAKKYVNTKPYYYYFHRGDSITTKPYAEKDLAYIDIYQKFYNEVV
KNYPDLKEVAFFRLAYAHFFILDKMLDDQYKQFEAYSQIHRFLKGHAFASIRNPFRKGRRISALALFINISLYRFLLLK
NIEKSKLLHZ

ATGAGAATCAAAGAGAAAACCAATAATATTAATGGAGGAATAAAAAATGTAAGTAAGCATTATGGTCATTCAATC
ATTCTCAAAGATATAAAATTTGCACTTAACAGAGGTGAACAAATTTGTTGGTCTACGAGGGAAGAAATGGAGTTGGTAAG
AGTACGTTGATGAAAATCTTTGTCAGAAATACGAGGACCTCAGGTAATATTATAAGCAGTGATAATGTTGGGT
ATTTAATCGAAGAACCAAAATTTATTTTATCTAAACAGGTTTAGAGAAATTTAAATATTTGTCAAATTTATATGG
TGTTGACTCAACTAAGAGAAAGATTTAGATGTTTGATCCAAGACTAGATTTGACTCAGTCTATTATAAAAAAGTA
AAGACCTATTCTTTGGGACAAAAACAAAAATAGCTTTGCTTCTGTAAGTCTCGTTACGGAACTGATATATTGATTA

TAGATGAACCGACTAATGGTTTAGATATTGAATCATCACAATAGTTTTAGCGGTTCTAAAAAATTAGCTTTACA
TGAAAAATGTGGGAATTTAATATCGAGTCATAAATTAGAAGACATTGAAGAAATTTGTGAGAGAGTTCTTTCTTG
GAGAACGGGCTTTGACATTTCAAAAAGTAGGAAAAGATAGTCATAATTTCTTTGTTGAGATAGCTTTTTCATCAG
CTACAGATAGAGACATTTTCATTACCAACAAGAATTTGGGATATTGTTTAG

MRIKEKTNNINGGIKNVSKHYGHSJLKDINFALNKEIVGLAGRNGVKGSTLMKILVQNNQPTSGNISSDNVGYLIEEP
KLFLSKTGLENLKYLNLVGVVDYNQERFRLIQELDLTQSINKVKVTSYSLGKQKLALLLTLVTEPDILILDEPTNGLDIE
SSQIVLAVLKKLALHENVGLISSHKLEDIEECERVLFLENGLLTFQKVGKDSHNFLEIAFSSATDRDIFITKQEFWDIVZ

ID25 1704bp

ATGACTGAATTAGATAAACGTCACCGCAGTAGCATTATGACAGCATGGTTAAATCACCTAACCGTGCTATGCTTC
GTGCGACTGGTATGACAGATAAGGACTTTGAAACATCGATTGTGGGAGTGATTTCGACTTGGGCGGAAAAATACAC
CATGTAACATTCACTTGCATGATTTGCGGAACTGGCTAAAGAAGGTGTCAAATCTGCAGGCGCTTGGCCTGTAC
AGTTTGGAAACATTACCGTAGCGGACGGGATCGCTATGGGAAACGCTGGTATGCGTTTCTCTCTAACATCTCGTGA
CATCATCGCGGACTCCATCGAGGCGGCTATGAGTGGTCACAACGTGGATGCCTTCGTCGCTATCGGTGGCTGTGA
CAAGAACATGCCTGGATCTATGATTGCTATTGCTAATATGGATATCCAGCTATTTTCGCTATGGTGGAACTATT
GCACCGGGAATCTTGATGGTAAAGATATCGACTTGGTTTCTGTCTTTGAAGGTATCGGAAAAATGGAACACCGT
GACATGACAGCTGAGGACGTGAAACGTCTTGAATGTAATGCCTGCCCTGGCCTGGTGGTTGTGGTGGTATGTAT
ACTGCTAATACCATGGCAACTGCTATCGAAGTTCTAGGGATGAGTTTGCAGGGTCTCCTCTCACCCAGCTGAAT
CAGCTGATAAGAAAGAAGATATCGAAGCAGCAGGACGTGCTGTTGTTAAGATGTTGGAACCTGGTCTCAAACCAT
CAGATATCTTGACTCGTGAAGCCTTTGAAGATGCTATCACTGTAACGATGGCTCTCGGTGGTTCTACAAACGCCAC
TCTTCACTTGGCTCGCCATGGCCATGCCGCAATGTTGACTTGTCACTTGAGGACTTCAATACGATTCAAGAACGT
GTGCTCACTTGGCCGACTTGAAACCATCTGGTCAATGTTCTTCAAGACCTCTACGAAGTCGGTGGTGTCCCTG
CGTTATGAAGTATTTGTTGGCAATGGTTTCTTACGGAGATCGCATCACATGTAAGTGGTAAAGTGTAGCTGA
AACTTGGCTGACTTTGCACTTGAAGTCTGACTCCAGGCCAAAAAGTTATCATGCCACTTGAAAAATCCAAACGTGCGGA
TGGTCCGCTTATCATCTTGAACGGGAACCTTGTCTGCTGACGGTGCAGTTGCCAAGGTATCAGGTGTTAAAGTGGCT
CGTCACGTTGGGCCAGCTAAGGTCTTTGACTCAGAAGAAGATGCGATTACGGCCGTTCTGACAGATGAAATCGTT
GATGGCGATGATAGTCGTTGTTCTGTTTGTGGACCTAAAGGTGGTCTGGTATGCTGAGATGCTATCACTTTCTTC
AATGATTGTTGGTAAAGGTGAGGAGATAAGTGGCCCTTTCAGCGGACGGACGTTTCTGTTGGTACTTATGGT
CTGGTTGTTGGACATATCGCTCCTGAAGCTCAGGATGGTGGACCAATTGCCTATCTCCGTACCGGCGATATCGTTA
CGTTGCCACCACTTTACAGCCGTTGCTCCTCGGTAATATGCCACATCGTATCATCTGCTTACCGCGGAGCCG
TGACAGACTTCTGGAATATGGACAAGTCAGGTAATAAATAA

MTELDKRHRSSIVDSMVKSPNRLRATGMTDKDFETSIIVGVISTWAENTPCNIHLHDFGKLAKGKVSAGAWPVQFG
TITVADGIAMGTPGMRFLTSRDIIADSIEAAMSGHNVDFAVIAIGGCDKNMPSMIAIANMDIPAIFYGGTIAPGNLDG
KDIDLVSFVFEIGIKWNHGDMTAEDVKRLECNACPGPGCGGMYTANTMATAEVLGMSLPSSSHPAESADKKEDIE
AAGRAVVKMLELGLKPSDILTREAFEDAITYTALGGSTNATLHLLAIAHAANVDLSLEDFTIQRVPHLADLKPSGQ
YVFQDLVEYVGGVPAVMKYLLANGFLHGDRTCTGKTVAENLADFADLTPGQKVMPLNPKRADGPLIILNGLNLPDG
AVAKVSGVKVRRHVGPAAKVFDEEDAIAQAVLTDEIVDGDVVVRFVGPKGPGMPPEMLSLSSMIVGKGQGDVKVALLT
DGRFSGGTGYGLVVGHIAPEAQDGGPIALYLRGTDIVTVDQDTKEISMAVSEEELEKRAETTLPLYSRGLVGLKYAHIVSS
ASRGAVTDFWNMDKSGKKZ

ID26 274bp

ATGTTATAATAAAAAATAAGAAATTTAAGGAGAAATACAATATGTCAATTTTTATTGGAGGAGCATGGCCATATGC
AAACGGTTTCGTTACATATTGGTCACGCGGAGCGCTTTTACCGGGGGATATTCTGCAAGATACTATCGTCAGAA
GGGAGAGGAAGTTTTATATGTTTCTGGAAGTGATTGTAATGGAACCCCTATTCTATCAGAGCTAAAAAGAAAA
TAAGTCTGTGAAAGAAATTGCTGATTTTATCATAAGGAATTTAATCCA

CYNKNKEFKENYMSIFIGGAWPYANGSLHIGHAAALLPGDILARYRQKGEEVLYVSGSDCNGTPISIRAKKENKSVK
EIADFYHKEFNP

ID28 1065bp

ATGACAACATTATTTTCAAAAATTAAGAAGTAACAGAACTTGCTGCAGTCTCAGGTCATGAAGCGCCTGTCCGT
GCTTATCTTCGTGAAAAGTTGACACCGCATGTGGATGAAGTGGTGACAGATGGCTTGGGTGGTATTTTTGGTATCA
AACATTACAGAAGCTGTGGATGCACCGCGCTCTTGGTCTCTCATATGGACGAAGTTGGTTTTATGGTCAGCGA
AATCAAGCCAGATGGTACCTTCGGTGTCTGAGAAATCGGTGGCTGGAACCCCATGGTGGTTAGCAGCCAACGTTT
CAAACTCTTGACTCGTGATGGTCATGAAATTCCTGTGATTTTCAAGTTCTGTTCTCCGCAATTGACTCGTGAAAG
GGGGGACCAACCATGCCAGCCATTGCCGATATCGTTTTGATGGTGGTTTTGCGGACAAGGCTGAGGCAGAAAGT
TTTGGCATCCGTCCTGGTGATACCATTTGACAGATAGTTCTGCAATTTTGACAGCCAATGAAAAAATATCATCT
CAAAAGCTTGGGATAACCGCTACGGTGTCTCATGGTAAGCGAGCTAGCTGAAGCTTTATCGGGTCAAAAACCTCG
GCAATGAACCTCTATCTGGGTTCTAACGTCCAAGAAGAAGTTGGTCTGCGTGGCGCTCATACCTCTACAACCAAGTT

5 TGACCCAGAAGTCTTCCTCGCAGTTGATTGCTCACCAGCAGGTGATGTCTACGGTGGTCAAGGCAAGATTGGAGA
TGGAACCTTGATTGCTTTCTATGATCCAGGTCACTTGCTTCTCCAGGGATGAAGGATTTCTTTTGACAAACGGCT
GAAGAAGCTGGTATCAAGTACCAATACTACTGTGGTAAAGGCGGAACAGATGCAGGTGCAGCTCATCTGAAAAAT
GGTGGTGTCCCATCAACAATATCGGTGTCTGCGCTCGTTATATCCATTCTCACCAAACCCCTCTATGCAATGGATG
ACTTCCTAGAAGCGCAAGCTTTCTTACAAGCCTTGGTGAAGAAATTGGATCGTTCAACGGTTGATTGATTAACA
TTATTAA

10 MTLFSKIKEVTELAASVSGHEAPVRAYLREKLTPHVDEVVTDGLGGIFGIKHSEAVDAPRVLVASHMDEVGFMVSEIKP
DGTFRVVEIGWNPVMVSSQRFKLLTRDGHEIPVISGSVPPHLTRGKGGTTPAJADIVFDGGFADKAEASFIRPGDT
IVPDSSAILTANEKNIISKAWDNRYGVLVSELAELSGQKLGNELYLSNVQEEVGLRGAHTSTTKFDPFVFLAVDCS
PAGDVYGGQKIGDGLIRFYDPGHLLLPKMDFLTTAEAGIKYQYYCGKGGTDAGAAHLKNGGVSTTIGVCARY
IHSHTLYAMDDFLEAQFLQALVKLDRSTVDLKHYZ

15 ID31 1182bp

ATGGAATTTTCTATGAAATCAGTCAAAGGACTACTCTTTATCATAGCTAGTTTTATCTTGACTCTTTGACTTGGAT
GAACACTTCTCCCAATTCTATGATTCCAGGACTAGCTTTAAACAGCCTATCTCTGACTTTTATCTAGCCACTCGT
CTCCCACTACTAGAAAGCTGGTTTACAGTTTGGAGAAGGTCTACACCGTCCACAAATTACAGCCTTTCTCTCAA
TCATCTCTAATCTTTATAACTTTAGTATGGGCGGTTTGTGGGGCTCTCGCTAGCTGCTCAGTTTGGCAATCTT
GCCATCTATATCTTTGCCAGCATCATCCTTGTGCGCTATTTAGGCAAATACATCCAATACGAAGCTTGGCGATGGA
TTCACCGCCTGGTTTACCTAGCCTATATTTAGGACTCTTTCACATCTACATGATAATGGGCAATCGTCTCCTTACA
TTTAACTCTTAAGTTTCTTGTGGTATGCTTGGCCTTTTAGGCTTACTAGCTGGTTTTATATCATTTTCTATAT
CAAAAGATTTCTTCCCTATCTAGGGAATAATACCATCTCAAACGCTTAAATCAGGATACTAGAGAAATTCAA
ATCCATCTTAGCAGACCTTTCAACTATCAATCAGGACAAATTTGCCTTTCTAAAGATTTTCCAAGAAGGCTTTGAAA
GTGCTCCGCATCCCTTTTCTATCTCAGGAGGTCTGGTCAAACCTTTACTTTACTGTTAAACTTCAGGCGACCA
TACCAAGAATATCTATGATAATCTTCAAGCCGGCAGCAAAGTAAACCTAGACAGAGCTTACGGACACATGATCAT
AGAAGAAAGGACGAGAAATCAGGTTTGGATTGCTGGAGGTATTGGGATCACCCTTCTCTTACATCCGTGA
ACATCCTATTTTAGATAAACAGGTTCACTTCTACTATAGCTTCCGTGGAGATGAAAATGCAGTCTACCTAGATTTA
CTCCGTAATCTATGCTCAGAAAAATCCTAATTTTGAATCCATCTAATCGACAGTACGAAAGACGGCTATCTTAAT
30 TTGAACAAAAAGAGTGCCCGAATGCAACCGTCTATATGTGTGGTCTATTTCTATGATGAAGGCACTTGCCA
AACAGATTAAGAAACAAATCCAAAAACAGAGCATATTTAC

35 MEFSMSVKGLLFIASFILTLTWMNTSPQFMIPGLALTSLSLTFILATRLPLESWFHSLEKVYTVHKFTAFLSIILLIFH
NFSMGGWLWGSRLAAQFNLAIYIFASIIIVAYLGYQYEAWRWIHRLVYLA YILGLFHIYIMMGNRLTFLNLSFLVGS
YALLGLLAGFYIIFLYOKISFPYLGKITHLKRNLNHDREIQIHLRPFNYQSQQFAFLKIFQEGFESAPHPFISISGGHQTLY
FTVKTSGDHTKNIYDNLQAGSKVTLDRAYGHMIEEGRENQVWIAAGGIGTTFISYIREHPILDKQVHFYYSFRGDENAV
YLDLLRNYAQKNPNFELHLIDSTKDGYNFEQKEVPEHATVYMCGPISMMKALAKQKKQNPKEHY

40 ID32 900bp

ATGACTTTTAAATCAGGCTTTGTAGCCATTTTAGGACGTCCTCAATGTTGGGAAGTCAACCTTTTAAATCAGCTTA
TGGGGCAAAAGATTGCCATCATGAGTGACAAAGCGCAGACAACGCGCAATAAAATCATGGGAATTTACACGACTG
ATAAGGAGCAAATTGTCTTTATCGACACACAGGGATTACAAAGCCTAAACAGCTCTCGGAGATTTCTATGGTTG
45 AGTCTGCCTACAGTACCTTTCGCGAAGTGGACACTGTTCTTTTCTATGGTGCCTGCTGATGAAGCGCGTGGTAAGGG
GGACGATATGATTATCGAGCGTCTCAAGGCTGCCAAGGTTCTGTGATTTTGGTGGTGAATAAAATCGATAAGGTC
CATCCAGACCAAGCTCTGTCTCAGATTGATGACTCCGTAATCAAATGGACTTTAAGGAAATGTTCCAATCTCAG
CCCTTCAGGGAATAACGTGTCTCGTCTAGTGGATATTTGAGTGAAAATCTGGATGAAGGTTTCCAATATTTCCC
GTCTGATCAAATCAGACCATCCAGAACGTTTCTTGGTTTCAGAAATGGTTCGCGAGAAAGTCTTGACACCTAACT
50 CGTGAAGAGATTCCGATTCTGTAGCAGTAGTTGTTGACTCTATGAAACGAGACGAAGAGACAGACAAGGTTTAC
ATCCGTGCAACCATCATGGTTCGAGCGGATAGCCAAAAAGGATTATCATCGGTAAAGGTGGCGCTATGCTTAAAG
AAAATCGGTAGCATGGCCCGTCTGTATCGAACTCATGCTAGGAGACAAGGTCTTCTAGAAACCTGGGTCAAG
GTCAAGAAAAACTGGCGGATAAAAAGCTAGATTTGGCTGACTTTGGCTATAATGAAAGAGAATACTAA

55 MTFKSGFVAILGRPNVGKSTFLNHVMGQKLAIMSDKAQTTNRNKIMGIYTTDKEQIVFIDTPGIHKPKTALGDFMVESAYS
TLREVDTVLFMVPADEARGKGDMMIERLKAAPVPLVVKIDKVPDQLLSQIDDFRNQMDKFEIVPISALQGNVVS
RLVDILSENLDGFFQYFSPDQITDHPERFLVSEMVREKVLHLTREIPHSVA VVVDMSMKRDEETDKVHIRATIMVERDSQ
KGIIIGKGGAMLKKIGSMARRDIELMLGDKVFLETWVKVKNWRDKKLDLADFGYNERYZ

60 ID33 855bp

CTGCTTCTGTTTTTACAGAAGGAGGACTTATGCCTGAATTACCTGAGGTTGAAACCGTTTGTCTGGCTTAGAAA
AATTGATTATAGGAAAGAAGATTTTCAGTATAGAAATTCGCTACCCCAAGATGATTAAAGACGGATTTGGAAGAGT
TTCAAAGGGAATTCCTAGTCAGATTATCGAGTCAATGGGACGTCGTGAAAAATATTTGCTTTTATCTGACAGA
65 CAAGGTCTTGATTTCCCATTTGCGGATGGAGGGCAAGTATTTTACTATCCAGACCAAGGACCTGAACGCAAGCAT
GCCCATGTTTTCTTCATTTTGAAGATGGTGGCAGCTTGTTTATGAGGATGTTTCGCAAGTTTGAACCATGGAAC

5 TCTTGGTGCCTGACCTTTTAGACGTCTACTTTATTTCTAAAAAATTAGGTCCTGAACCAAGCGAAACAAGACTTTGA
TTTACAGGCTTTTCAATCTGCCCTTGCCAAAGTCCAAAAAGCCTATCAATCCCATCTCCTAGACCAGACCTTGGTA
GCTGGACTTGGCAATATCTATGTGGATGAGGTTCTCTGGCGAGCTCAGGTTTCATCCAGCTAGACCTTCCAGACTT
TGACAGCAGAAGAAGCGACTGCCATTCATGACCAGACCATTTGCTGTTTTGGGCCAGGCTGTTGAAAAAGGTGGCT
CCACCATTGGACTTATACCAATGCCTTTGGGGAAGATGGAAAGCATGCAGGACTTTCATCAGGCTATGATAAGA
CTGGTCAAGAATGTGTACGCTGTGGTACCATTCATTGAGAAAAATCAACTAGGCGGACGTGGAACCCACTTTTGTCC
AAACTGTCAAAGGAGGACTGA

10 MLLVFTEGGMLPELPEVETVCRGLEKLIQKISSIEIRPKMIKTDLEEFQRELPSQIESMGRRGKYLLFYLTDKVLISHL
RMEGYFYYPDQGPERRKHAHVFFHFEDGGTLVYEDVRKFGTMELLVPDLLDVYFISKKLGPPESEQDFDLQVFSALA
KSKKPIKSHLLDQTLVAGLNIYVDEVLRWRAQVHPARPSQTLTAEETAJHDQTLAVLQGAVEKGGSTIRTYTNAFGED
GSMQDFHQVYDKTGQECVRCGTIIEKIQLGGRGTHFCPNCQRRDZ

15 ID34 633bp

TTGTCCAAACTGTCAAAGGAGGGACTGATGGGAAAAATCATCGGAATCACTGGGGGAATTGCCTCTGGTAAGTCA
ACTGTGACAAATTTTCTAAGACAGCAAGGCTTTCAAGTAGTGGATGCCGACGCAGTCGTCACCAACTACAGAAA
CCTGGTGGTCGTCTGTTTGAGGCTCTAGTACAGCACTTTGGGCAAGAAATCATTCTTGAAAACGGAGAATCAATC
GCCCTCTCCTAGCTAGTCTCATCTTTTCAAATCCTGATGAACGAGAATGGTCTAAGCAAATTCAGGGGAGATTAT
20 CCGTGAGGAACTGGCTACTTTGAGAGAACAGTTGGCTCAGACAGAAAGAGATTTTCTTCATGGATATTCCTCTACTT
TTTGACAGGACTACAGCGATTGGTTTGTGAGACTTGGTGGTCTATGTGGACCGAGATGCCAAAGTGGAAACGC
TTAATGAAAAGGGACCAAGTTGTCCAAAGATGAAGCTGAGTCTCGTCTGGCAGCCCAAGTGGCCTTTAGAAAAAAG
AAAGATTTGGCCAGCCAGGTTCTTGATAATAATGGCAATCAGAACCAGCTTCTTAATCAAGTGCATATCCTTCTTG
AGGGAGGTAGGCAAGATGACAGAGATTA

25 MSKLSKEGLMGKIIIGTGGIASGKSTVTNFLRQQGFQVVDADAVVHQLQKPGGRLFEALVQHFGQEIIENGENLNRPLL
ASLIFSNPDEREWSKIQGHEIREELATLREQLAQTEIFFMDIPLLFEQDYSDFWFAETWLVYVDRDAQVERLMKRDQLS
KDEAESRLAAQWPLEKKDLASQVLDNNGNQNLNQHVLLEGGRRQDDRZ

30 ID35 1269bp

TTGATAATAATGGCAATCAGAACCAGCTTCTTAATCAAGTGCATATCCTTCTTGAGGGAGGTAGGCAAGATGACA
GAGATTAAGTGAAGGATAATCTGCGCATTGCCTGGTTTGGTAATTTCTGACAGGAGCCAGTATTTCTTTGGTTG
TACCTTTATGCCCATCTTCGTGGAAAAATCTAGGTGATGGGAGTCAGCAAGTCGCTTTTATGACAGGCTTAGCAAT
35 TCTGTCTCTGCTATTTCCGCGCGCTCTTTCTCCTATTTGGGGTATTCTTGCTGACAAATACGGCCGAAAACCCA
TGATGATTCCGGCAGGTCCTTGTCTATGACTATCACTATGGGAGGCTTGGCCTTTGTCCCAAATATCTATTGGTTAAT
CTTTCTTCGTTTACTAAACGGTGTATTTGCAAGTGTGTTCTTAATGCAACGGCACTGATAGCCAGTCAGGTTCCA
AAGGAGAAATCAGGCTCTGCCTTAGGTACTTTGTCTACAGGCGTAGTTGCAAGTACTCTAACTGCTCCCTTTATTG
GTGGCTTTATCGCAGAAATATTGGCAATTCGTACAGTTTCTTACTGGTTGGTAGTTTCTATTCTTAGCTGCTATT
40 TGACTATTTGCTTTATCAAGGAAGATTTTCAACCAGTAGCCAAGGAAAAGGCTATTCCAAACAAAGGAATATTATAC
CTCGGTTAAATATCCCTATCTTTTGTCTCAATCTCTTTTAAACAGTTTGTCTATCCAAATTTTCACTCAATCGATTG
GCCCTATTTTGGCTCTTATGTACCGCACTTAGGGCAGACAGAAATCTTCTTTTGTCTCTGGTTTGATTGTGTCC
AGTATGGGCTTTTCCAGCATGATGAGTGCAGGAGTCATGGGCAAGCTAGGTGACAAGGTGGGCAATCATCGTCTC
45 TTGGTTGTGCGCCAGTTTATTCAGTCATCATCTATCTCTCTGTGCCAATGCCTCTAGCCCCCTTCAACTAGGACT
CTATCGTTTCTCTTTGGATTGGGAACCGGTGCCTTGATTCCCGGGGTTAATGCCCTACTCAGCAAAATGACTCCC
AAAGCCGGCATTTCGAGGGTCTTTGCCTTCAATCAGGTATTCTTTATCTGGGAGGTGTTGTTGGTCCCATTGGCAG
GTTCTGCAGTAGCAGGTCAATTTGGCTACCATGCTGCTTTTATGCGACAAGCCTTTGTGTTGCCTTTAGTTGTCTC
TTTAACTGATTCAATTTGCAACATTATTAAGAAAGTAAAGGAAATCTAG

50 MIIMAIRTSFLIKISFLREVGMTEINWKDNLRIA WFGNFLT GASISLVVFPMPFIVENLGVSQQVAFYAGLAISVS AIS
AALFSPWIGILADKYGRKPMIRAGLAMTITMGLAFVFNLYWLFRLNLNGVFAGFVFNATALLASQVPKEKSGSALG
TLSTGVVAGTLTGPFIGGFIAELFGIRTVFLLVGSFLFLAAILTICFIEDFQPVAKEKAIPTELFTSVKYPYLLNLFLTS
FVIQFSAQSIGPILALYVRDLGQTENLLFVSGLVSSMGFSSMMSAGVMGKLGDKVGNHRLLVVAQFYSVIYLLCANAS
55 SPLQLGLYRFLGLGTGALIPGVNALLSKMTPKAGISRVFAFNQVFFYLGGVVGPMAAGSAVAGQFGYHAFYATSLCV
AFSCLFNLIQFRTLKVKEIZ

ID36 1311bp

ATGGCCCTACCAACTATTGCCATTGTAGGACGTCCCAATGTTGGGAAATCAACCCTATTTAATCGGATCGCTGGTG
AGCGAATCTCCATTGTAGAAGATGTCGAAGGAGTGACACGTGACCGTATTTATGCAACGGGTGAGTGGCTCAATC
60 GTTCTTTTATGACATGATTGATACAGGAGGAATTGATGATGTCGATGCTCCTTTTATGGAACAAATCAAGCACCAGGC
AGAAATGGCATGGAAGAAGCAGATGTTATCGTTTTTGTCTGTCTGGTAAGGAAGGAATTACTGATGCAGACGA
ATACGTAGCTCGTAAGCTTTATAAGACCCACAAACAGTTATCCTCGCAGTCAACAAGGTGGACAACCCCTGAGAT
GAGAAATGATATATATGATTCTATGCTCTCGGTTTGGGTGAACCATTCCTATCTCATCTGTCATGGAATCGGT
65 ACAGGGGATGTGCTAGATGCGATCGTAGAAAATCTTCAAATGAATATGAGGAAGAAAAATCCAGATGTCATTAAAG
TTTAGCTTGATTGGTCGTCTAACGTTGGAAAAATCAAGCTTGATCAATGCTATCTTGGGAGAAGACCGTGTTATTG

CTAGTCCTGTTGCTGGAACTCGTGATGCCATTGATACCCACTTTACAGATACAGATGGTCAAGAGTTTACCAT
GATTGATACGGCTGGTATGCGTAAGTCTGGTAAGGTTTATGAAAATACTGAGAAATACTCTGTTATGCGTGCCATG
CGTGCTATTGACCGTTTACAGATGTGGTCTTGATGGTCATCAATGCGGAAGAAGGCATTGAGTACGACAAGCGT
5 ATGCGCAGGATTTGCCATGAAGCTGGTAAAGGGATGATTATCGTGGTCAACAAGTGGGATACGCTTGAAAAAGAT
AACCACACTATGAAAACTGGGAAGAAGATATCCGTGAGCAGTTCCAATACCTGCCTTACGCACCGATTATCTTT
GTATCAGCTTTAACCAAGCAACGTCTCCACAACTTCTGAGATGATTAAAGCAAATCAGCGAAAGTCAAAATACA
CGTATTCCATCAGCTGTCTTGAACGATGTCTATGATGGATGCCATTGCCATCAACCCAAACCGACAGACAAAGGA
10 AAACGTCTCAAGATTTTCTATGCGACCAAGTGGCAACCAAAACCAACCTTTGTATCTTTGTCAATGAAGAAAG
AACTCATGCACTTTTCTTACCTGCGTTTCTTGGAAATCAAATCCGCAAGGCCTTTGTTTGTAGGGAACACCGAT
TCATCTCATCGCAAGAAACGCAATAA

MALPTIAIVGRPNVKGSTLFRNIAGERISTVEDVEGVTRDRIYATGEWLNRSFSMIDTGGIDDVDAPFMEQIKHQAEIAM
EEADVIVFVVSQKEGTTDADEYVARKLYKTHKPVLAVNKVDNPEMRNDIYDFYALGLGEPLPISSVHIGTGDVLDAL
15 VENLPNEYEEENPDVIFSLIGRPNVGKSSLINAILGEDRVIASPVAGTTRDAIDTHTDGTGQEFMTIDTGMKRSKGV
YENTEKYSVMRAMRAIDRSVVLMVINAEEGIREYDKRIAGFAHEAGKGMIIIVNKNWDTLEKDNHMTKNWEEDIREQ
FQYLPYAPIIFVSALTQRLHKLPEMIKQISESQNTRIPSAVLNDVIMDAIAINPTPTDKGKRLKIFYATQVATKPPTFVIFV
NEEELMHFSYLRFLNQIRKAFVFEGTPIHLIARKRKZ

ID37 714bp

20 ATGACAGAAACCATTAATTTGATGAAGGCTCATCTTCAGTGCGCAGGTTTAAAGAGCAAGAAATCCCCAAGTA
GACTTAAATGAGATTTTGACAGCAGCCAGATGGCATCATCTTGGAGAATTTCCAATCCTACTCTGTGATTGTGG
TACGAAGTCAAGAGAAGAAAGATGCCTTGTATGAATTGGTACCTCAAGAAGCCATTGCGCAGTCTGCTGTTTCTCT
TCTCTTTGTGCGAGATTTGAACCGAGCAGAAAAGGGAGCCGACTTCATACCGACACCTTCAACCCCAAGGTGT
25 GGAAGGTCTCTTGATTAGTTCGGTTCGATGCACTCTTGGTGGACAAAACGCCTTGTGGCAGCTGAAAGCTTGGGC
TATGGTGGTGTGATTATCGGTTTGGTTCGATACAGTCTGAAGAAGTGGCAGAGCTCTTAACTACCTGACTACA
CCTATTCTGTCTTTGGGATGGCACTGGGTGTGCCAAATCAACATCATGATATGAAACCGAGACTGCCACTAGAGA
ATGTTGTCTTTGAGGAAGAATACCAAGAACAGTCAACTGAGGCAATCCAAGCTTATGACCGTGTTCAGGCTGACT
30 ATGCTGGGGCGCGTGGCACCACAAGCTGGAGTCAGCGCTAGCAGAACAGTTTGGTCAAGCTGAACCAAGCTCAA
CTAGAAAAATCTTGACAGAGAAATTTATTGTAG

MTETIKLKAHTSVRRFKEQEIQVLDNEILTAAQMASSWKNFQSYSVTVVRSQEKDALYELVPQEAIRQSAVFLLFV
GDLNRAEKGARLHTDTFQPGVEGLLISSVDAALAGQNALLAESLGYGGVIIGLVRYKSEEVAFNLNPDYTSYVFG
35 MALGVPNQHDMKPRPLENVVFEEYQEQSTEAJQAYDRVQADYAGARATTSWSQRLAEQFGQAEPSSTRKNLEQK
KLLZ

ID38 729bp

40 ATGACAGAAATTAGACTAGAGCAGTCAGTTATGCCTATGGTCAGGAGAGGATTTTAGAGGATATCAACCTACAG
GTGACTTCAGGGCAAGTGGTTTCCATCCTAGGCCCAAGTGGTGTGGAAAGACCACCTCTTAACTAATCGCTG
GGATTTAGAAGTTCAGTCAGGAGAAATGTCCTTGATGGTGAAGAAAATCCCAAGGGGCGGTGAGTTATATGT
TGCAAAAGGATCTGCTCTTGGAGCACAAGACGGTGTGGAAATATCATTCTGCCCTCTTGATTCAAAAGGTGG
ATAAGGCAGAAAGCTATTTCCGAGCGGATAAAATCTTGGCAGCTTCCAGCTGACAGCTGTAAGAGACAAGTATC
45 CTCATGAACCTTAGCGGTGGGATGCGCCAGCGTGTAGCCTTACTCCGACCTACCTTTTGGGCACAAGCTCTTTCT
CTTAGATGAGGCCTTAGCGCCTTGGATGAGATGGAAGATGGAAGTCCAGCTTGGTATCTTGAGATTCACAA
GCAGTTGCAGCTAACAACCTGATCATACGCATAGTATTGAGGAGGCCCTCAATCTCAGCGACCGTATCTATATC
TTGAAAAATCGCCCTGGGCAGATTGTTTCAAGAAATTAAGTAGATTGGTCTGAAGATGAGGACAAGGAAGTCCAA
AAGATTGCCTACAAACGTCAAAATTTGGCGGAATTAGGCTTAGATAAGTAG

50 MTEIRLEHVSAYGQERILEDINLQVTSGEVVSILGPSGVGKTTLFLNIAGILEVQSGRIVLDGEENPKGRVSYMLQKDLL
LEHKTVLGNILPLLIQKVDKAEISRADKILATFQLTAVRDKYPHELSSGMRQRVALLRITYLFGHLFLLDFAFSALDE
MTKMELHAWYLEIHKQLQLTTLITHSIEEALNLSDRYILKNRPGQIVSEIKLDWSEDEKQKIAKYRQILAEGLDK
Z

ID39 2433bp

55 ATGAACTATTCAAAAGCATTGAATGAATGTATCGAAAGTGCCTACATGGTTGCTGGACATTTTGGAGCTCGTTATC
TAGAGTCGTGGCACTTGTGATTGCCATGTCTAATCACAGTTATAGTGTAGCAGGGGCACTTTAAATGATTATCC
GTATGAGATGGACCGTTTGAAGAAGGTGGCTTTGGAACCTGACTGAAACGGACTATAGCCAGGATGAAACCTTTAC
60 GGAATTTGCCGTTCTCCCGTCTTTGACAGGTTCTTTTGTATGAAGCAGAGTATGTAGCGTCAGTGGTCCATGTAAG
GTACTAGGGACAGAGCAGTCCTCTATGCGATTTTGCATGATAGCAATGCCTTGGCGACTCGTATCTTGGAGAGG
GCTGGTTTTTCTTATGAAGACAAGAAAGATCAGGTCAAGATTGCTGCTCTTCGTCGAAATTTAGAAGAACGGGCA
GGCTGGACTCGTGAAGATCTCAAGGCTTTACGCCAACGCCATCGTACAGTAGCTGACAAGCAAAATCTATGGCC
AATATGATGGGCATGCCGAGACTCCTAGTGGTGTCTCGAGGATTATACGCATGATTTGACAGAGCAAGCGCGT
65 TCTGGCAAGTTAGAACCAGTCATCGGTGGGACAAGGAAATCTCACGTATGATTCAAATCTTGAGCCGGAAGACT

5 AAGAACAACCCCTGCTTGGTTGGGGATGCTGGTGTGCGGAAAAACAGCTCTGGCGCTTGGTCTTGGCCAGCGTATTG
CTAGTGGTGACGTGCGTTCGCGAAATGGCTAAGATGCGCGTGTAGAACTTGATTTGATGAATGCTGTCAGGGA
CAGCTTCCGTGGTGACTTTGAAGAACGCATGAATAATATCATCAAGGATATTGAAGAAGATGGCCAAAGTCATCC
10 TCTTTATCGATGAACTCCACACCATCATGGGTTCTGGTAGCGGGATTGATTCGACTCTGGATGCGGCCAATATCTT
GAAACCAGCCTTGGCGCGTGGAACCTTTGAGAACGGTTGGTGCCACTACTCAGGAAGAATATCAAAAAACATATCGA
AAAAGATGCGGCACCTTCTCGTCGTTTCGCTAAAGTGACGATTGAAGAACCAGTGTGGCAGATAGTATGACTAT
TTTACAAGGTTTGAAGGCGACTTATGAGAAACATCACCGTGTACAAATCACAGATGAAGCGGTTGAAACAGCGGT
15 TAAGATGGCTCATCGTTATTTAACAGTCGTCACTTGCCAGACTCTGCTATCGATCTCTGGATGAGGCGGCAGCA
ACAGTGCAAAATAAGGCAAGCATGTAAAGCAGACGATTGAGTTGAGTCCAGCTGACAAGGCCCTGATGGAT
GGCAAGTGGAAAACAGGCAGCCAGCTAATCGCAAAAGAAGAGGAAGTACCTGTCTACAAAGACTTGGTGACAGA
GTCTGATATTTTGACCACCTTGAGTCGCTTGTGAGGAATCCCACTTCAAAAACTGACTCAAACGGATGCTAAGAAG
TATTTAAATCTTGAAGCAGAACTCCATAAACGGGTTATCGGTCAAGATCAAGCTGTTTCAAGCATTAGCCGTGCCA
20 TTGCGCCGAACCAAGTCAGGGATTCCGAGTCATAAGCGTCCGATTGGTTCCTTTATGTTCTAGGCTACAGGTGT
CGGAAAAAATTGAATTAGCCAAGGCTCTGGCAGAAGTCTTTTACGACGAATCAGCCCTTATCCGCTTGTATG
15 AGTGAGTATATGGAGAAATTTGCAGTAGTCGTCTCAACGGAGCTCTCCAGGCTATGTAGGATATGAAGAAGGT
GGGAGTTGACAGAGAAGGTTTCGCAATAAACCTTATCCGTTCTCTCTTTGATGAGGTAGAGAAGGCCACCCA
GATATCTTTAATGTTCTCTGACAGGTTCTGGATGACGGTGTCTTGACAGATAGCAAGGGACGCAAGGTCGATTTT
CAATACCATTTATCATTATGACATCGAATCTAGTGCGACTGCCCTTCGTGATGATAAGACTGTTGGTTTGGGCG
25 TAAGGATATTCGTTTGGCAGGAAAATATGGAAGAACGCGATGTTTGAAGAAGTGAAGAAAGCTTATAGACCGGA
ATTCATCAACCGTATTGATGAGAAGGTGGTCTTCCATAGCCTATCTAGTGATCATATGCAGGAAGTGGTGAAGATT
ATGGTCAAGCCTTTAGTGGCAAGTTGACTGAAAAGGCGATTGACTTGAATTACAAGCTTCAAGCTCTGAAATGT
TAGCAAAATCAAGGATATGACCCAGAGATGGGAGCTCGCCCACTTCGAGAACCCTGCAAAACAGAGTGGAGGAC
AAGTTGGCAGAACTTCTTCTCAAGGGAGATTTAGTGGCAGGCAGCACCTAAGATTGGTGTCAAAGCAGGCCAG
25 TTAATAATTTGATATTGCATAA

30 MNYSKALNECIESAYMVAGHFGARYLESWHLIAMSNSHSYVAGATLNDYPYEMDRLEEVLELTETDYSQDETTFE
LPFSRRLQVLFDEAEYVASVVHAKVLGTEHVLYAILHDSNALATRLERAGFSYEDKKDQVKIAALRRNLEERAGWTR
EDLKALRQRHRTVADKQNSMANMMGMPQTPSGGLEDYTHDLTEQARSQKLEPVIGRDKESIRMIQILSRKTKNNPVLV
GDAGVGKTALALGLAQRIASGDVPAEMAKMRVLELDLMNVVAGTRFRGDFEERMNNIKDIEEDGGVILFIDELHTIM
35 GSGSGIDSTLDAANILKPALARGTLRTVGATTQEEYQKHIEKDAALSRRFAKVITIEEPSVADSMITLQGLKATYEXHHRV
QITDEAVETAVKMAHRYLTSRHLPSAIDLDEAAATVQNKAKHVKADDSDLSPADKALMDGKWQKQAAQLIAKEEV
PVYKDLVTESDILTSLRSGIPVQKLTQTDACKYLNLEAELHKRVIGQDQAVSSISRAIRRNQSGIRSHKRPIGFSFMLGP
TGVGKTELAKALAEVLFDDESALIRFDMSEYMEKFAASRLNGAPPGYVGYEEGGELTEKVRNKPYSVLLFDEVEKAHP
35 DIFNVLLQVLDGVLTDKGRKRVDFSNTHIMTSNLGATALRDDKTVGFGAKDIRFDQENMEKRMFEELKKAYPEFIN
RIDEKVVHSLSSDHMQEVVKIMVKPLVASLLEKIDKLQASALKLLANQGYDPEMGARPLRRLTQTEVEDKLAELL
LKGDVLVAGSTLKIGVKAGQLKFDIAZ

40 ID40 1008bp

45 ATGAAGAAAACATGGAAAGTGTTTTTAACGCTTGTAAACAGCTCTTGTAGCTGTTGTGCTTGTGGCCTGTGGTCAAG
GAACTGCTTCTAAAGACAACAAAGAGGCAGAACTTAAAGAGTTGACTTTATCCTAGACTGGACACCAAAATACCA
ACCAACAGGGCTTTATGTTGCCAAGGAAAAGGTTATTTCAAAGAAGCTGGAGTGGATGTTGATTTGAAATTCG
45 CACCAGAAGAAAGTTCTTCTGACTTGGTTATCAACGGAAGGCACCAATTGCAAGTGTATTTCAAAGACTACATGGC
TAAGAAATTTGAAAAAGGAGCAGGAATCACTGCCGTTGCAGCTATTGTTGAACACAATACATCAGGAATCATCTC
TCGTAATCTGATAATGTAAGCAGTCCAAAAGACTTGGTTGGTAAGAAATATGGGACATGGAATGACCCAACTGA
ACTTGCTATGTTGAAAACCTTGGTAGAATCTCAAGGTGGAGACTTTGAGAAGGTTGAAAAAGTACCAAAATAACGA
CTCAAACTCAATCACACCGATTGCCAATGGCGCTTTTGATACTGCTTGGATTTACTACGGTTGGGATGGTATCCTT
50 GCTAAATCTCAAGGTGTAGATGCTAACTTCATGACTTGAAAGACTATGTCAAGGAGTTTGACTACTATTACCCAG
TTATCATCGCAAAACACGACTATCTGAAAGATAACAAAGAAGAGCTCGCAAAGTCATCAAGCCATCAAAAAA
GGCTACCAATATGCCATGGAACATCCAGAAGAAGCTGCAGATATTCTCATCAAGAATGCACCTGAACCTCAAGGAA
AAACGTGACTTTGTATCGAATCTCAAAAATCTGTCAAAAGAATACGCAAGCGACAAGGAAAAATGGGGTCAA
TTTGACGAGCTCGCTGGAATGCTTTCTACAAATGGGATAAAGAAAATGGTATCCTTAAAGAAGACTTGACAGAC
AAAGGCTTACCAACGAATTTGTGAAATAA

55 MKKTWKVFLTLVTALVAVVLVACGQGTASKDNKEAELKKVDFILDWTPNTNHTGLYVAKEKGYFKEAGVDVCLKLP
PESSSDLVINGKAPFAVYFQDYMAKKLEKGAGITAVAIVHNTSGIISRKSDNVSSPKDLVGKKYGTWNPTELAML
KTLVESQGGDFEKVEKVPNNDSNSITPIANGVFDTAWIYYGWDGILAKSQGV DANFMYLKDYYKEFDYSPVIANND
60 YLKDNEEARKVIQAIKKGYQYAMEHPPEEAADILIKNAPELKEKRFVIESQKYLSEYASDKEKWGFDAARWNAFY
KWDKENGILKEDLTDKGFTNEFVKZ

65 ID41 762bp

TTGATGAGAACTTGAGAAGTATACTGAGACGACACATTAGTCTATTGGGCTTCTCGGAGTATTGTCAATCTGGC
AGTTAGCAGGTTTTCTAAACTTCTCCCAAGTTATCCTGCCGACACCTCTGAAATCTCCAGCCCTTTGTTCTG
GACAGAGAATTTCTCTGGCACCATAGCTGGGCGACCTTGAGAGTGGCTTTACTGGGGCTGATTTGGGAGTTTGTGA

- 5 TTGCTGTCTTATGGCTGTGCTCATGGATAGTTTGACTTGGCTCAATGACCTGATTTACCTATGATGGTGGTCATT
CAGACCAATCCGACCATTTGCCATAGCTCCTATCCTGGTCTTGTGGCTAGGTTATGGGATTTGCCCAAGATTGTCT
TGATTATCTTAACGACAACCTTTCCCATCATCGTTAGTATTTTGGACGGTTTTAGGCATTGGCACAAGGATATGCT
GACCTTGTTTAGTCTGATGCGGGCCAAGCCTTGGCAAACTCTGTGGCATTTTAAAATCCCAAGTAGCCTGCCTTAC
TTTTATGCAGGTCTGAGGGTCAGTGTCTCCTACGCCCTTATCACAACCTGTGGTATCTGAGTGGTTGGGAGGTTTTG
AAGGCTTGGTGTTTATATGATTCACTTAAAAAAGTGTTCAGTATGATACCATGTTTGCCATTATTATCTGGTG
TCGATTATCAGTCTTTTGGGTATGAAGCTGGTCGATATCAGTGAAAAATATGTGATTAATGGAACGTTCTGAT
- 10 MMRNLRSLRRHISLLGFLGVLISWQLAGFLKLLPKFILPTPLEILQPFVRDREFLWHHSWATLRLVALLGLLGLVLIACLM
AVLMDSLTWLNDLYPMVMVVIQTIPTIAIPILVLWGLYGLPKIVLILTTTFPIVSILDGFRHCDKMDLTLFSLMRAP
WQILWHFKIPVSLPYFYAGLRVSVSYAFITTVVSEWLGGEGLGVYMIQSKLFLQYDTMFJHILVSIISLLGMKLVDSIE
KYVIKWKRSZ
- 15 ID42 372bp
TTGATTTTAACTCTATTTGCTGTATGATAAGGGAAAAGAAAGGGGACAGATATGGCTTTTACCAATACCCACA
TGGCATCTGCTAGTTTGGTATTGTTACCAGCTTGCTGATGACATCATTGACTCTTTTGGTATATCATCGACCAT
TTCTTAAAAAATGTCTTTGAATTGGAAGAAGAACTCGAGTTTCAATTGCTTAATAACCAAGGAAAGATTACCTTCC
ACTTTTCAAGTCAACACCTCCCTACAGCCATTGATTTTGACTTTAACCATCCTTTTCGACCCTCGTTATCCCCAAGA
GTACTGGTTTTAGACATGGACGGTAGAGAAACTATCCTCCTCCAGAAGAAAATGACCTATTTTAA
- 20 MIFNPICCMIREKKGDRDMAFTNTHMRSASFGIVTSLPDDIIDSFWYIHDHFLKNVFELEEELEFQLLNNQGKITFHSSQ
HLPTAIDFDNFHFPDPRYPRLVLVLDMDGREILLPEENDLFZ
- 25 ID43 1569bp
ACAGCGGTGTCATTCTATCTATTTTAAAGAAAAGTAATAATCAATTGTTAAAAATAGTAAAAAATTTGGAGGTTCTG
ATGAAAATTTTGTTCCTAATGAGGTATTCAGTATTCGTAATAAAGGTGGGGAAGTCTGCTCGGTACTATTGGCAA
TTTCAATTTTGGGAAGCCAAGGTATTTATCGGATGAAGTTGTTACTAGTTCTTCCCGATGGCTACAAAAGAGTC
TTCTAATGCAATTACTAATGATTTAGATAATTCACCAACTGTTAATCAGAATCGTTCTGCTGAAATGATTGCCTCT
AATTCAACCACTAATGGTTTAGATAATTCGTTAAGTGTTAATAGCATCAGCTCTAATGGTACTATTGTTCCAATT
CACAATTAGACAACAGAACAGTTGAATCTACAGTAACATCTACTAATGAAAATAAGAGTTATAAGGAAGATGTTA
TAAGTGACAGAATTATCAAAAAAGAAATTTGAAGATACTGCTTTAAGTGTAAGAGATTATGGTGCAGTAGGTGATG
GGATTCTATGATGATCGACAAGCAATTCAAGATGCAATAGATGCTGCAGCTCAAGGGCTAGGTGGAGGAAATGTAT
35 ATTTTCTGAAGGAACTTATTTAGTAAAAGAAATTTTAAAAAGTCATACACACTTGAATGAATGAGAA
AGCTACAATTCTAAATGGTATAAATATTAAGAATCACCTTCCATTGTTTTATGACAGGTTTATTACGGATGAT
GGTGCGCAAGTAGAATGGGGCCCAACAGAGATATTAGTTATTCTGGTGGTACGATTGATATGAACGGGTGCTTTG
AATGAAGAAGGAACTAAAGCAAAAAATCTACCACTTATAAATCTTCAGGTGCAATTTGCTATTGGGAATTCAAAT
AACGTAATATAAAAAATGTAACTTCAAGGATAGTTATCAAGGGCATGCTATTCAAATTCAGGTTTCGAAAAAT
40 GTATTAGTTGATAATTCTCGTTTCTTGGGCAAGCCTTACCCAAAACGATGAAGGATGGGCAATCATAAGTAAGG
AGAGCATTGAGATTGAACCATTAAGTAAAAAGGTTTCTTATGCCTTGAATGATGATGGGAAAAAATCTGAAA
ATGTGACTATTCAAAATTCCTATTTTGGCAAAAGTGATAAATCTGGGGAATTAGTAACAGCAATTGGCACACACTA
TCAAACATTGTCGACACAGAACCCCTCTAATATTAATAATCAAAATAATCAATTTGATAACATGATGTATGACGGT
GTACGTTTTACAGGATTCAGTATGATTAATCAAAGGAAATCGCTTTGATAAGAAAGTTAAAGGAGAGAGTGTA
45 CATTATCGAGAAAGCGGAGCAGCTTAGTAAATGCTTATAGCTATAAAAACTAAAGACCTATTAGATTTAAAT
AAACAGGTGGTTATCGCCGAAAAATATATTTAATATTGCCGATCCTAAAACAAAAGCGATACGAGTTGCAAAAGAT
AGTGACAGAAATGTTAGGAAAAGTATCAGATATTACTGTAACAAAAATGTAATTAATAAATCTAAGGAAAAA
GAACAACCAATATTGAATTATTACGAGTTAGTGATAATTTAGTAGTCTCAGAGAATAGT
- 50 QRCHSIYFKKSNNQLLKIVKKLEVLKMYFVPNEVFSIRKLKVGTCVLLAISILGSQGILSDEVVTSPPMATKESSNAITN
DLDNSPTVQNRSIAEMIASNSTTNGLDNSLSVNSISSNGTIRSNSQLDNRTVESTVTSTNENKSYKEDVISDRIKKEFEDT
ALSVKDYGAVDGIHDDRQAIQDAIDAAAQGLGGGNYFPEGTYLVKEIVFLKSHTHLELNEKATILNGINIKNHPISIVF
MTGLFTDDGAQVEWGPTEDISYSGGTIDMNGALNEEGTKAKNPLINSSGAFAGNSNNVTIKNVTFKDSYQGHAIQIA
55 GSKNVLVDNSRFLGQALPKTMKDGQIISKESIQIEPLTRKGFPYALNDDGKKSENVTIQNSYFGKSDKSGELVTAIGHY
QTLSTQNPSNIQNNHFDNMMYAGVRFGTFTDVLKGNRFDKKVKGESVHYRESGAALVNAYSYKNTKDLLDLNKQ
VVAENIFNIADPKTKAIRVAKDSAELGKVSIDITVTKNVINNNSKETEOPNIELLRVSDNLVVSNS
- 60 ID44 324bp
GTGATGAAAGAACTCAGCTATTTAAAGGTGTTCTTGAAGGTTGTGCTTGGATATGATTGGTCAAAAAGAGCGG
TATGTTATGAGTTGGTTGAGCTTTGCGAGAGGCTGGATTTGATACTATCGTTCCAGGAATTTATCCTTTGTT
GCAAAAGTTAGAAAAAATCAATGGATAAGAGGCGACATGCGCCCGTCGCCAGATGGTCCAGATCGGAAGTATTT
TTCATTAATGAAAGAAAGGAGAAGAGCGTGTCTCAGTCTTTTGGCAACATGGGACGATTGAGTCAAAAAGTAGA
65 AGGGATTAAGAAATGGGGTTAA

MMKETQLLKGVLEGCVLDMIGQKERYGYELVQTLREAGFDITVPGTIYPLLQKLEKNQWIRGDMRPSDPGPRKYFSL
MKEGEERVSVFWQWDDLQKVEGIKNGGZ

ID45 816bp

5 ATGAAGAAAATGAAGTATTACGAAGAAAACAGCGCTTTGCTACATGAGTTTCTGAGGAGAATCAAAAGTATTTT
GAGGAGTTGTGGGAAAGTTTAACTCTGCTGGATTCTCTATGATGAAGACTATCTCAGAGAGCAGATCTATTGTA
10 TGATGCTAGATTTCTCAGAAGCAGAACGAGATGGCATGAGTGCAGAGGATTATCTAGGTAAAGATCCTAAAAAAA
TAATGAAAGAGATTCTCAAGGGAGCACCTCGCAGTTCTATCAAAGAGTCCCTTTTGACGCCAATTCTGTCTGGC
GGTATTACGTTATTATCAACTACTAAGTGATTTTCTAAAGGTCCTCTTAAACAGTCAATTTGCTCACATTTTATG
GGCAACTTCTTATTTTCTGATTGGATTGGACTGTGGCCACAATTTTACGAAGAAGTTTAGTCCAAGATTCTCCT
AAAATGAAAATTGGCACTTACATTGTTGTTGGGACTATAGTTCTTCTAGTTGTTTATAGGATATGTAGGAATGGCAA
GCTTCATACAAGAAGGAGCCTTTATATTCCGGCTCCCTGGGATAGTTTGTCTGTTTACGATTTCGCTAGTTATC
15 GGTATTGGAATTGGAAGAAGCGGTCTTTCGTCATTTGTCAGTATGATTATGCCCATCTGTGGTGGGTTCTCT
GCTCGTTATTATGAGTGGATGGGAATTTCAAATGTTTTCTTACAAAAGTTATTCCTTAGCTGTCTCTTATTG
GAATCTTTGTCTTTCGTTGGGTTAAGAAGATAAAATGGAGTGAAGTATAG

20 MKMKMYEETSALLHEFSEENQKYFEELWESFNLAGFLYDEDYLRQIYLMMLDFSEAERDGMSEAEDYLGNPKKIM
KEILKGAPRSSIKESLLTPILVLAFLRYQLLSDFSKGPLLTVNLLTFLGQLLIFLIGFGLVATILRRSLVQDSPKMKIGTYI
VVGITVLLVVLGYVGMASFIQGAIFYAPWDSLSVFTISLVIGIWNWKEAVFRPFVSMIAHLVVGSLRLRYEWMGISN
VFLTKVIPLAVLFIGIFVLRFGFKKIKWSEVZ

ID46 348bp

25 CTGTTTTTTTATTATACTCAATGAAAATCAAAGAGCAAACTAGGAAGCTAGCCGAGGTTGCTCAAAACACTGTT
TTGAGGTTGTAGACGAACTGACGAAGTCAGCTCAAAACATGTTTTGAGGTTGTAGATGAACTGACGAAGTCA
GCTCAAAACACTGTTTTGAGGTTGTAGATGAACTGACGAAGTCAGCTCAAAACACTGTTTTGAGGTTGTAGATG
AAACTGACGAAGTCAGCTCAAAACATGTTTTGAGGTTGTAGATGAAACTGACGAAGTCAGTAACCATACATACG
30 GTAGGGCAGCGCTGACGTGGTTTGAAGAGATTTCGAAGAGTATTAA

MFFYLYSMKIKEQTRKLAAGCSKHCFEVVDETDEVSSKHVFEVVDETDEVSSKHCFEVVDETDEVSSKHCFEVVDET
EVSSKHVFEVVDETDEVSNHTYGRATLTWFEEIFEYZ

ID47 1260bp

35 ATGCAGAATCTGAAATTTGCCTTTTCATCTATCATGGCTCACAAAGATGCGTTCTTTGCTTACTATGATTGGGATTAT
TATCGGTGTTTCATCAGTTGTTGTGATTATGGCTTTGGGTGATTCCCTATCTCGTCAAGTCAATAAAGATATGACTA
AATCTCAGAAAAATATTAGCGTCTTTTCTCTCTCTAAAAAAGTAAAGACGGGTCTTTTACTCAGAAACAATCAGC
40 TTTTACGGTTTCTGGAAAGGAAGGGAAGTCTCTGTTGAACCGCCAAAACCGCAAGAATCCTGGGTCCAAGAGGC
AGCTAACTGAAGGGAGTGGATAGTTACTATGTAACCAATTCAACGAATGCCATCTTGACCTATCAAGATAAAAA
GGTTGAGAATGCTAATTTGACAGGTGGAACAGAACTTACATGGACGCTGTTAAGAAATGAAATTATTGCAGGTG
TAGTCTGAGAGAGCAAGATTTCAAAGAGTTTGAAGTGTCTTTTGTAGATGAGGAATTGTCCATTAGTTTATTT
GAATCTCTCAAGAGGCTATTAACAAGGTTGTAGAAGTCAATGGATTAGTTACCGGGTCAATGGGTTTATACTA
45 GTCCGGAGGCTAAAAGATCAAAAATATATGGGTTTGGTGGCTTGCTATTACTACCAATATCTCCCTTGCTGCGAA
TTTTAATGTAGATGAAATAGCTAATATTGCTTTTCGAGTGAATGATACAGTTTAACCCCACTCTGGGTCCAGAA
CTGGCAGCAAAAATGACAGAGCTTGACGGCTTACAACAGGGAGAATACCAGGTGGCAGATGAGTCCGTTGTATT
GCAGAAATCAACAATCGTTTAGTTTTATGACGACGATTATTAGTTCCATCGCAGGGATTCTCTCTTTGTTGGAG
GAACTGGTGTGATGAACATCATGCTGGTTTCGGTGACAGAGCGCACTCGTGAGATTGGTCTTCGTAAGGCTTTGGG
50 TGCAACACGTGCCAATATTTAATTCAGTTTGTGATTGAATCCATGATTTTGACCTTGTAGGTGGCTTAATTGGCT
TGACAATTGCAAGTGGTTAACTGCCTTAGCAGGTTTGTACTGCAAGGTTTAAAGAAAGGTATAGAAGTTGGAGT
ATCAATCCAGTCGCCCTATTTAGTCTTGACGTTTCGGCTAGTGTGGTATGATTTTTGGAGTCTGCCAGCCAAC
AAGGCATCGAAACTTGATCCAATTGAAGCCCTTCGTTATGAATGA

55 MQNLKFAFSSIMAHKMRSLTMTGIIIGVSSVVVIMALGDSLSRQVKNKDMTKSQKNISVFFSPKKSXKDGSTQKQSAFTVS
GKEEEVPEPPKQESVWQEAALKGVDSYYVTNSTNAULTYQDKKVENANLTGGNRTYMDAVKNEIAGRSRLEQDF
KEFASVILLDEELSISLFESPQEAINKVVEVNGFSYRVIGVYTSPEAKRSKIYFGGLPITTNISLAANFNVDIANTVFRVN
DTSLTPTLGPGLARKMTLAGLQQGEYQVADESVFVFAEQSFMTTISSIAGISLVGGTGVMNIMLVSVTERTREIG
LRKALGATRANILIQFLIESMILTLLGGLIGLTIASGLTALAGLLLQGLIEGIEVGVSIPVALFSLAVSASVGMIFGVLPANK
60 ASKLDPIEALRYEZ

ID48 705bp

65 CTGATGAAGCAACTAATTAGTCTAAAAATATCTTCAGAAGTTACCGTAATGGTGACCAAGAACTGCAGGTTCTC
AAAAATATCAATCTAGAAGTGAATGAGGGTGAAATTTGTAGCCATCATGGGACCATCTGGGTCTGGTAAGTCCACT
CTGATGAATACGATTGGCATGTTGGATACCAACCAGTGGAGAATATTATCTTGAAGGTCAAGAAGTGGCTGGG

5 CTTGGTGAAAAACAAGCTAAGGTCCGTAACCAACAAATCGGTTTTGTCTTTACGACAGTTCTTCTTCTATCGA
AGCTCAATGCTCTGCAAAATGTAGAAATGCCCTTGATTTACGCAGGAGTTTCGTCTTCAAAACGTCGCAAGTTGGC
TGAGGAATATTTAGACAAGGTTGAATTGACAGAACGTAGTCACCATTTACCTTCAGAATTATCTGGTGGTCAAAA
GCAACGTGTAGCCATTGCGCGTGCCTTGGTAAACAATCCTTCTATTATCCTAGCGGATGAACCGACAGGAGCCTTG
GATACCAAAACAGGTAACCAAAATTATGCAATTATTGGTTGATTTGAATAAAGAAAGGAAAAACCATTATCATGGTA
ACGCATGAGCCTGAGATTGCTGCCTATGCCAAACGTCAGATTGTCATTGCGGATGGGGTCATTTCTGCTGACAGTG
CTCAGTTAGGAAAGGAGGAAAACTAA

10 MMKQLISLKNIFRSYRNGDQELQVLKNINLEVNEGEFVAIMGPSGSGKSTLMNTIGMLDTPSTGEYYLEGQEVAGLGEK
QLAKVRNQIQGFVQFFLLSKLNALQNVLEPLIYAGVSSSKRRKLAEEYLDKVELTERSHHLPSELSSGGQKQVRVAIARA
LVNPNPILADEPTGALDTKGNQIMQLLVLDLNKEGKTIIMVTHEPEIAAYAKRQIVIRDGVISSDAQLGKEENZ

ID49 1200bp

15 ATGAAGAAAAAGAAATGGTAAAGCTAAAAAGTGGCAACTGTATGCAGCAATCGGTGCTGCGAGTGTAGTTGTATTG
GGTGCTGGGGGGATTTTACTCTTTAGACAACCTTCTCAGACTGCTCTAAAAGATGAGCCTACTCATCTTGTGTTG
CCAAGGAAGGAAGCGTGGCCTCTCTGTTTTATTGTCAGGGACAGTAACAGCAAAAAATGAACAAATATGTTTTATT
TTGATGCTAGTAAGGGTGATTTAGATGAAATCCTTGTCTTCTGTTGGCGATAAGGTGAGCGAAGGGCAGGCTTTAGT
20 CAAGTACAGTAGTTCAGAAGCGCAGGCGGCTATGATTCAGCTAGTCGAGCAGTAGCTAGGGCAGATCGTCATAT
CAATGAACTCAATCAAGCACGAAATGAAGCCGCTTCAGCTCCGGCTCCACAGTTACCAGCGCCAGTAGGAGGAGA
AGATGCAACGGTGCAAGGCCAACTCCAGTGGCTGGAAATCTGTTGCTTCTATTGACGCTCAATTGGGTGATGCC
CGTGATGCGCGTGACAGATGCTGCGGCGCAATTAAGCAAGGCTCAAAAGTCAATTGGATGCAACAACTGTTCTCAGT
ACCTTAGAGGGAAGTGTGGTGAAGTCAATAGCAATGTTTCTAAATCTCCAACAGGGGCGAGTCAAGTTATGGTT
CATATTGTCAGCAATGAAAATTTACAAGTCAAGGGAGAATTGTCTGAGTACAATCTAGCCAACTTTCTGTAGGTC
25 AAGAAGTAAGCTTTACTTCTAAAGTGTATCCTGATAAAAAATGGACTGGGAAATTAAGCTATATTCTGACTATCC
TAAAAACAATGGTGAAGCAGTAGTCCAGCAGCCGGAATAATACAGGTTCTAAATACCTTTATACTATTGATGT
GACAGGCGAGGTTGGTGATTTGAAACAAGGTTTTCTGTCAACATTGAGGTTAAAGCAAACTAAGGCTATTCTT
GTTCTGTAGCAGTCTAGTAATGGATGATAGTAAAAATTATGTCTGGATTGTGGATGAACAACAAAAGGCTAAAG
30 AAAGTTGAGGTTTCATTGGGAAATGCTGACGCAGAAAAATCAAGAAATCACTTCTGGTTTAAACGAACGGTGCTAAG
GTCATCAGTAATCCAACATCTTCTTGAAGAAGGAAAAAGAGGTGAAGGCTGATGAAGCAACTAATTAG

35 MKKKNGKAKKWQLYAAIGAASVVVLGAGGILLFRQPSQTALKDEPHTLVVAKESVASSVLLSGVTYAKNEQYVYFD
ASKGDLDEILVSVGDKVSEQALVKYSSSEAQAA YDSASRAVARADRHINELNQARNEAASAPAPQLPAPVGGEDATV
QSPTPVAGNSVASIDAQLGDARDARADAAAQLSKAQSLDATTVLSTLEGTVEVNSNVSKSPTGASQVMVHVSNEN
LQVKGELSEYNLANLSVGOEVSFTSKVYPDKKWTGKLSYSDYPKNNGEASPAAGNNTGSKYPYTIDVTGEVGLKQ
GFSVNIEVSKTKAILVPVSSLVMDDSKNYVWVDEQQAkakKVEVSLGNADAENQEITSGLTNGAKVISNPTSSLEEGKE
VKADEATNZ

ID50 759bp

40 ATGTCACGTAACCATTATCGCTGGTAACTGGAAAAATGAACAAAAATCCAGAAGAAGCTAAAGCATTCTGTTGAA
GCAGTTGCATCAAAAACCTTCTTCATCAGATCTTGTGAAGCAGGTATCGCTGCTCCAGCTCTTGATTTGACAACTG
TTCTTGCTGTTGCAAAAGGCTCAAACTTAAAGTTGCTGCTCAAACTGCTACTTTGAAAAATGCAGGTGCTTTTAC
45 TGGTGAAACTAGCCCAAGTTTTGAAAGAAATCGGTACTGACTACGTTGTTATCGGTCACTCAGAACGCCGTGA
CTACTTCCATGAAACTGATGAAGATATCAACAAAAAGCAAAAGCAATCTTTGCGAACGGTATGCTTCCAATCAT
CTGTTGTGGTGAACTCACTTGAACCTTACGAAGCTGGTAAAGCTGCTGAATTCTGAGGTGCTCAAGTATCTGCTGCA
TTGGCTGGATTGACTGCTGAACAAGTTGCTGCCTCAGTTATCGCTTATGAGCCAATCTGGGCTATCGGTACTGGTA
AATCAGCTTCAACAGACGATGCACAAAAATGTGTAAGTTGTTGCTGACGTTGTAGCTGCTGACTTTGGTCAAG
AAGTCGACAGACAAAGTTCTGTTCAATACGGTGGTTCTGTTAAACCTGAAAAATGTTGCTTCATACATGGCTTGCCC
-50 AGACGTTGACGGTGCCCTTGTAGGTGGTGGCTCACTTGAAGCTGAAAGCTTCTTGGCTTGTGCTTGTGAAAA
TAA

55 MSRKPIAGNWKMNKNPEEAKAFVEAVASKLPSSDLVEAGIAAPALDLTTVLAVAKGSLKVAQAQNCYFENAGFTG
ETSPQVLKEIGDYVIGHSERRDYFHETDEINKKAKAIFANGMLPIICCGESLETYEAGKAAEFVGAQVSAALAGLTA
EQVAASVIA YEPIWAI GTGKSASQDDAQKMCKVVRDVVAADFGQEVADKVRVQYGGSVKPNVASVMACPDVDGAL
VGGASLEAESFLALLDFVKZ

ID51 1473bp

60 TTGAAAAACAAAATTGGATTAGCAAGTATCTGTTTACTAGGCTTGGCAACTAGTCATGTCGCTGCAAAATGAAACTG
AAGTAGCAAAAACCTTCGACAGGATACAACGACAGCTTCAAGTAGTTCAGAGCAAAATCAGTCTTCTAATAAAACGC
AAACGAGCGCAGAAGTACAGACTAATGCTGCTGCCCACTGGGATGGGGATTATTATGTAAGGATGATGGTTCTA
AAGCTCAAGTGAAATGGATTTTGAACACTACTATAAGGCTTGGTTTTATATTAATTCAGATGGTCTGTTACTCGCA
GAATGAATGGCATGGAATTAACCTGAAATCAGGTGGATATATGGCCCAAAACGAGTGGATCTATGACAGTAA
65 TTACAAGAGTTGTTTTATCTCAAGTCAGATGGGGCTTATGCTCATCAAGAAATGGCAATTGATTGGAATAAGTGG

TACTACTTCAAGAAGTGGGGTTACATGGCTAAAAGCCAATGGCAAGGAAGTTATTTCTTGAATGGTCAAGGAGCT
ATGATGCAAAATGAATGGCTCTATGATCCAGCCTATTCTGCTTATTTTTATCTAAAATCCGATGGAACCTATGCTA
ACCAAGAGTGGCAAAAAGTGGGCGGCAAAATGGTACTATTTCAAGAAGTGGGGCTATATGGCTCGGAATGAGTGGC
AAGGCAACTACTATTTGACTGGAAGTGGTGGCCATGGCGACTGACGAAGTGATTATGGATGGTACTCGCTATATCTT
5 TGGCGCCTCTGGTGAGCTCAAAGAAAAAAGATTGAATGTCGGCTGGGTTACAGAGATGGTAAGCGCTATTT
CTTTAATAATAGAGAAGAACAGTGGGAACCGAACATGCTAAGAAAGTCATTGATATTAGTGAGCACAATGGTCG
TATCAATTGATTGAAAAAGTTATTGATGAGAACGAAGTGGATGGTTCATTGTTCTAGTTATAGCGGTAA
AGAAGACAAGGAATTGGCGCATAACATTAAGGAGTTAAACCGTCTGGGAATTCCTTATGGTGTCTATCTCTATAC
10 CTATGTGAAAAATGAGACCGTGTGAGAGTGACGCTAAACAGACCATTGAACCTATAAAGAAATACAATATGAAC
CTGTCTTACCTCTATCTATTATGATGTTGAGAATTGGGAATATGTAATAAGAGCAAGAGAGCTCCAAGTGATACA
GGCACTTGGGTTAAAAATCATCAACAAGTACATGGACACGATGAAGCAGGCGGGTTATCAAAATGTGTATGTCTAT
AGCTATCGTAGTTTATTACAGACGCTTAAAAACCCAGATATTTTAAACATGTAACTGGGTAGCGGCCTATA
CGAATGCTTTAGAATGGGAAAACCTCATTATTCAGGAAAAAAGGTTGGCAATATACCTCTCTGAATACATGA
15 AAGGAATCCAAGGCGCGTAGATGTCAGCGTTTGGTATTAA

MKTKIGLASICLLGLATSHVAANETEVAKTSQDTTASSSSEQNQSSNKTQTSAEVQTNAAAHWDGDYVYKDDGSKAQ
SEWIFDNYKAWFYINSDGRYSQNEWHGNYLKSQGYMAQNEWIYDSNYKSWFYLKSDGAYAHQEWQLIGNKWYY
FKKWGYMAKSQWQGSYFLNGOGAMMONEWLYDPAYSAFYLKSDGTANQEWQKVGKWWYFKKWGYMARNE
20 WQGNYYLTGSGAMATDEVIMDGTRYIFAASGELKEKKDLNVGWVHRDGKRYFFNNREEQVGEHAKKVIDISEHNG
RINDWKKVIDENEVDGVVRLGYSGKEDKELAHNIKELNRLGIPYGVLYTYAENETDAESDAKQTELKKNYNNLSY
PIYDVENWEYVYVNSKRAPSDGTWVKIINKYMDTMKQAGYQNVVYYSRLLQTRLKHPDILKHVNWVAAVTNAL
EWENPHYSKKGWQYTSSEYMKGIQGRVDVSVWYZ

ID52 774bp

ATGAAAAAATTTGCCAACCTTATCTGGGACTGGTCTTCTGGTCTCTACCTGCCTATCTTTTACTTGATTGGCTA
TGCTTTAATGCTGGTGATGATATGAATAGCTTTACAGGTTTTAGCTGGACTCAGTTTGAAACCATGTTTGGAGAT
GGGAGACTCATGCTGATTGCTGCTCAGACATTTTCTGGCCTTCTATCAGCCTTGATAGCGACCATATCGGGA
30 CTTTGGTGCCATTACATCTACCACTCTCGTAAGAAATACCAAGAGCCCTTCTATCACTCAATAATATCCTCAT
GGTTGGCGCCTGACGTTATGATTGGTGCTAGCTTCTGATTCTCTTACCAACTCAAGTTTTCACTTGGCTTTTGA
CCGTCTATCTAGTCAGTGGCCTTCTCCATTCTATCGTGGTCTTGATGGTCTTGCTCGACTCAAGGAAATGAA
TGGCGACATGATTCTATGCGGCCTATGACTTGGGAGCTAGTCAATTCAGATGTTCAAGGAAATCATGCTTCTTAC
CTGACTCCGTCTATCATTACTGGTTATTTATGCGCTTACCTATTCGTTAGATGACTTTGCCGTGACCTTCTTTGT
35 AACAGGAAATGGCTTTTCAACCTATCAGTCGAGATTTACTCTCGTCTCGCAAGGGGATTTCCTTAGAAATCAAT
GCCCTGTCTGCTCTAGTCTTCTCTTATGATTATCTAGTTGTAGGTTATTACTTTATCTCTCGTGAGAAGGAGGA
GCAAGCATGA

MKKFANLYLGLVFLVLYLPIFYLIGYAFNAGDDMNSFTGFSWTHFETMFGDGRMLMLLAQTFFLAFLSALIATTIGTGA
IYIYQSRKKYQEAFLSLNNILMVAPDVIMIGASFLILFTQLKFLSLGFLTVLSSHVAFSIPVVLMLVPLRKEMNGDMIHAA
40 YDLGASQFQMFKEIMLPYLTPSITGYFMAFTYSLDDFAVTFVTGNGFSTLSVEIYSRARKGISLEINALSALVFLFSIILVV
GYFISREKEEQAZ

ID59 1071bp

ATGAAAAAATCTATTCATTTTATAGCAGGAATTGCAGCGATTATCCTTGTCTTGTGGGAATTGCGACTCATTAG
ATAGTAAAAATCAATAGTCGAGATAGTCAAAAATTTGGTTATCTATAACTGGGAGACTATATCGATCCTGAACTCTT
GACTCAGTTTACAGAAGAAACAGGAATCAAGTTCAGTACGAGACTTTTACTCCAACGAAGCCATGTACACTAA
GATAAAGCAGGGTGGAACGACCTACGATATTGCCATTCCAAGTGAATACATGATTAACAAGATGAAGGACGAAG
50 ACCTCTTGGTCCGCTTGATTATTCAAAAATGAAGGAATCGAAAATATCGGACCAGAGTTTCTCAACCACTCCTT
TGACCCAGGTAATAAATTCTCCATCCCTTACTTCTGGGGAACCTTAGGAATTGTCTACAACGAAACCATGGTAGAT
GAAGCGCCTGAGCATTGGGATGACCTTTGGAAGCCGAGTATAAGAATTCTATCATGCTCTTTGATGGGGCGCGT
GAGGTGCTGGGACTAGGACTCAATCCCTCGGCTACAGCCTCAACTCCAAGGATCTGCAGCAGTTGGAAGAGACA
GTGGATAAGCTCTACAACTGACTCCAATATCAAGGCTATCGTTGCGGACGAGATGAAGGGCTATATGATTTCAG
AATAATGTTGCAATCGCGTGACCTTCTCTGGTGAAGCCAGCCAAATGTTAGAAAAAATGAAAAATCTACGTTAT
55 GTGGTACCGACAGAGGCCAGCAATCTTTGGTTTGACAATATGGTCAATCCAAAACAGTTAAAAACCAAACTCA
GCCTATGCCTTTATCAACTTTATGTTGAAACCTGAAAAATGCTCTCCAAATGCGGAGATGTCGGCTATTCAACAC
CAAACCTACCAGCGAAGGAATTGCTCCAGAGGAAACAAAGGAAGATAAGGCCTTCTATCCCGATGTTGAAACCA
TGAAACACCTAGAAGTTTATGAGAAATTTGACCATAAATGGACAGGGAATATAGCGACCTCTTCTACAGTTTA
60 AAATGTATCGGAAGTAG

MKKIYSFLAGIAAJILVLWGIATHLDSKINSRDSQKLVIYNWGDYIDPELLTQFTEETGIQVQYETFDSDNEAMYTKIKQGG
TTYDIAIPSEYMINKMKDDELLVPLDYSKIEGIEINIGPEFLNQSFDPGNKFSIPYFWGTLGIVYNETMVDPEHWDDLW
KPEYKNSIMLFDGAREVLGLGLNSLGYSLNSKDLQLEETVDKLYKLTPNKAIVADEMKGYMIQNNVIAIGVTFSGEAS
65 QMLEKNENLRVVPTEASNLWFDNMVIPKTVKNQNSAYAFINFLKPENALQNAEYVGYSTPNLPAKELLPEETKED
KAFYDPDVETMKHLEVYEFKDHKWTGKYSDFLQFKMYRKZ

ID61 1851bp

5 ATGAATAAAAACTAACAGATTATGTGATTGATCTGGTGGAAATTTTAAATAAACAAACAAAAGCAGGTTTTCTGG
GGAATATTTGATATTTTCAGTATGGTGGTTTCCATCATTGTATCTTATATTTTATTTTATGGGCTGATTAATCCAGC
ACCTGTTGACTACATTATCTATACGAGTTTGGCCTTCTGTTCTATCAATTGATGATTGGTTTTTGGGGGTGAACG
CGAGCATTAGTCGTTACAGCAAGATTACGGATTTCATGAAAACTTTTTTGGTGTGACTGCTAGCAGTGTCTTGTG
ATATAGTATCTGTTATGCCTTCTTGCCACTCTTCTCCATCCGTTTCATCAITCTCTTTATCTTGTGAGTACCTTCTT
10 GATTTTATTGCCACGGATTACTTGGCAGTTAATCTACTCCAGACGCAAAAAAGGTAGTGGTGTGGAGAACCCG
TCGGACCTTCTGATTGGTGCCGGTGTGGTGGGGCTCTTTTATGGATAAGTTACCAACATCCAACAGTGAATTA
GAAGTGGTCGGTATTTTGGATAAGGATTCTAAGAAAAAGGGTCAAAAACTTGGTGGTATTCCTGTTTTGGGCTCTT
ATGACAATCTGCCTGAATTAGCCAAACGCCATCAAATCGAGCGTGTATCGTTGCGATTCCGTCGCTGGATCCGTC
AGAATATGAGCGTATCTTGAGATGTGTAATAAGCTGGGTGTCAAATGTTACAAGATGCCTAAGGTTGAAACTGT
15 TGTTCAGGGCTTACCAAGCAGGTAAGCTTCCAAAAATGATATTACGGACCTTTGGGTGCTCAGGAAATC
CGTCTTGACGAATCGCGTCTGGGTGCAGAACTGACAGGTAAGACCATCTTAGTCACAGGAGCTGGAGGTTCAATC
GGTTCGAAATCTGTCGTCAAGTTAGTCGCTTCAATCCTGAACGCATTGCTTGTCTGGTCATGGGAAAACTCAA
TCTACCTTGTATCATGAATTGATTCTGAAGTTCCAAGGATTGATTATGTACCTGTGATTGCGGACATTCAAGA
CTATGATCGTTTGTGCAAGTCTTTGAGCAGTACAAACCTGCTATTGTTTATCATGCGGCAGCCCAAGCATGTT
20 CCTATGATGGAGCGCAATCCAAAAAGAGCTTCAAAAAAATATCCGTGGAACCTTCAATGTTGCTAAGGCTGTT
GATGAAGCTAAAGTGTCTAAGATGGTTATGATTTGCAGAGATAAGGCAGTCAATCCACCAAAATGTTATGGGAGCA
ACCAAGCGCGTGGCGGAGTTGATTGCTACTGGCTTAAACCAAGTAGCCAATCAACCTACTGTGCAGTTCGTTTTG
GGAATGTTCTTGGTAGCCGTGGTATGTCATTCCAGTCTTTGAACGTGAGATTGCTGAAGGTGGGCTGTAACGGT
GACAGACTTCCGTATGACCCGTACTTTATGACCATTCAGAAAGCTAGCCGTCTGGTTATCCATGCTGGTGCTTAT
25 GCCAAAGATGGGGAAGTCTTTATCCTTGATATGGGCAAAACAGTCAAGATTTATGACTTGGCCAAGAAGATGGTG
CTTCTAAGTGGCCACACTGAAAGTGAATTCGAATCGTTGAAGTTGAATCCGCCAGGTGAAAACTCTACGAA
GAACTCTTGGTATCAACCGAACTCGTTGATAATCAAGTTATGGATAAGATTTTCTTGGTAAGGTTAATGTCATGC
CTTTAGAATCCATCAATCAAAAGATTGGAGAGTTCGCCACTCTCAGTGGAGATGAGTTGAAGCAAGCTATTATCG
CCTTTGCTAATCAACAACCCACATTGAATAA

30 MNKKLTDYVIDLVEILNKQKQVFWGIFDIFSMVSVIIVSYLYGLINPAPVDYIITYTSLAFLFYQLMIGFWGLNASISRY
SKITDFMKIFFGVTAASSVLSYSICYAFLPLFSIRFIILLSTFLILLPRITWQLIYSRRKKGSGDGEHRRFTFLIGAGDGGALF
MDSYQHPTSELELVGILDKDSKKKGQKLGIPVLGSYDNLPELAKRHQIERVIVAPSLDPSEYERILQMCNKLGVKCYK
MPKVETVVQGLHQAGTFQKIDITDLLGRQIEIRLDESRLGAELTGKILVTGAGGSIGSEICRQVSFRNPERIVLLGHGEN
35 SYLVYHELIRKFQGDYYPVVIADIQDYDRLLQVFEQYKPAIVYHAAAHKHVPMERNPKEAFKNIRGTYNVAKAVD
EAKVSKMVMISTDKAVNPPNVMGATKRVAAELIVTFGNQRSQSTYCAVRFGNVLSRGSVIPVFERQIAEGGPVTVDFR
MTRYFMTIPEASRLVIHAGAYAKDGEVFILDMGPKPKIYDLAKKMVLLSGHTESEIPIVEVGIRPGEKLYEELLVSTELV
DNQVMDKIFVGKVNVMPLSINQKIGEFRTLSGDELKQAIHAFANQTHIEZ

ID101 1338bp

40 ATGATTGAACCTTTATGATAGTTACAGTCAAGAAAGTCGAGATTTACATGAAAGTCTAGTCGCTACTGGTCTTTCTC
AAGTTGGAGTGGTCATCGATGAGATGGTTTTCTGCCTGATGGTCTGCTTTCTCCTTTTACCTATTATCTAGGTTAC
GAGGATGGAAAAACCTCTCTATTTTAAATCAAGTTCCCGTTTTCAGATTTTTGGGAAATTTTAGGAGATAATCAGCTCG
CTTGATTGAAGATGTGACGCAGGAGAGGGCTGTCTTTCATTATGCTGATGGAATGCAGGCTCGCTTGGTTAAACA
45 GGTAGACTGGAAAGACCTAGAAGGTGAGTACGTGAGTTGACCACTACAATCGCTTCGGAGCTTGTGTTGTCTAC
AACGACTTATAGCGCAGATAGCGAGCCGATTATGACAGTTTACCAAGATGTCAATGGTCAACAAGTTTACTGGA
AAACCATGTGACGGGTGATATCTTATTGACTTTGCCAGGTCAGTCCATGCGTTACTTTGCAAAATAAGTTGAATTT
ATCACCTTCTTTTGCAAGATTTGGAAATAGATACCAGTCAGCTTATCTTAACTACTAGCGACTCTTTCTTGGT
50 TTCTTCCATCATCCAGATAAATCTGGCTCGGATGCTTGGTATGGCAGGAACCTCTCTATGATGCCATCCAGGT
AATATGCAGTTGATTTTGGAAAGTGATAATGTGCGTACTAAGAAGATCATCTTCCAAATAAGGCGACTTATGAG
CGCGCTTTAGAGTTAACTGACGAGAAATACCATGATCAGTTTGTGCACTTGGGTTATCATTACGAGTTCAACCGTG
ATAATTTCTAAGACGAGATGCCTTAATCTTGACCAATTCAGATCAGATTGAGCAAGTAGAAGCAATCGCAGGAG
CCTTGCCTGATGTCACTTTCCGTATTGACGCGGTGACAGAGATGTCTTAAAGCTCTTAGACATGCTTTGCTATCCT
55 AATGTGGCCCTTTACCAGAACGCTAGTCCACAGAAGATTGAGGAGCTGTATCAACTGTGGATATTTACTTGGATA
TAAACCACAGTAATGAGTTGCTACAGGCAGTGGTCAAGGCTTTGAGCACAATCTCTTATTCTTGGCTTAAATCA
GACGGTGCAACAATAGACTTTATATCGCTCCAGACCATCTATTGAAAGTAGTGAAGTTGCTGCTTTGGTTGAGACC
ATTAATTTGGCCCTTTTTCAGATGTTGATCAATGCGTCAGGCACTTGGCAACAAGGCCAACATGCAAAATATGTTG
ACTTGGTGAGATATCAGGAAACCATGCAAACTGTTTTAGGAGGCTAA

60 MIELYDSYSQESRDLHESLVATGLSQLGVVIDADGFLPDGLLSPFTYYLYGYEDGKPLYFNQVPVSDFWELGDNQSACIE
DVTQERAVIHYADGMQARLVKQVDWKDLEGRVRQVDHYNRFGACFATTTYSADSEPIMTVYQDVNGQVLLNHHV
TGDILLTLPQGSMMRYFANKVEFITFLQDLEIDTSQILFNLTATPFLVSFHHDPKSGSDVLVWQEPYDAIPGNMQLILES
65 DNVRTKIIPNKATYERALELTDEKYHDFVHLGYHYQFKRDNFLRRDALILTNSDQIEQVEALAGALPDVTFRIAAVT
EMSSKLLDMLCYPNVALYQNASPQKIQLYQLSDIYLDINHNSNELLQAVRQAFEHNLILGFNQTVHNRLYIAPDHLFE
SSEVAALVETIKLALSDVDQMRQALGKQGHANYVDLVRYQETMQTVLGGZ

ID102 1512bp

5 ATGACAAATTTACAATATAAATTTAGGAATTGGTTGGGCTAGTAGCGGTGTTGAATACGCTCAAGCCTATCGTGCTG
 GTGTTTTTCGGAAATTAATCTGTCTCTAAGTTTATCTTTACAGATATGATTTTAGCCGATAATATTCAGCACTTA
 ACAGCCAAATATTGGTTTTGATGATAATCAGGTTATCTGGCTTTATAATCATTTACAGATATCAAAATTCACCTTA
 CTAGCGTGACAGTGGATGATGCTCTGGCTTACTTTGGTGGTGAAGAAAGTCACAGAGAAAAAATGGCAAGTTT
 TACGTGATATCTTTTTGACCAAGATAAGTTTGTAACCTGTTATTTGGTTGATGAGAACAAAGGACTTGGTTCAACA
 10 TGCCGAGTATGTTTTAAGGGAACCTGATTGCGAAGGATTACTTTCTTATACGCGTTATTGTAGCGAGTATTTT
 GCTCCCAAGGACAATGTTGCAGTCTTATACCAACGAACTTTTATAATGAAGACGGGACTCCAGTCTATGATATCT
 TGATGAATCAAGGGAAGGAAGAAGTTTATCATTTCAAGGATAAGATTTTCTATGGAAGCAAGCTTTTGTGCGTG
 CCTTTATGAAATCTTTGAATTTGAATAAGTCTGATTTGGTCATTCTCGATAGGGAGACAGGTATTGGACAGGTTGT
 GTTTGAGGAAGCACAGACAGCACATCTAGCGGTAGTTGTTTCATGCGGAGCATTATAGTGAAATGCTACAAATGA
 15 GGACTATATCCTTTGGAATAACTATTATGACTATCAGTTTACCAATGCAGATAAGGTTGACTTCTTATCGTGTCT
 ACTGATAGACAAAATGAAGTTCTACAAGAGCAATTTGCCAAATATACTCAGCATCAGCCAAAGATTGTTACCATT
 CCTGTAGGCAGTATTGATTCCTTGACAGATTCAAGTCAAGGGCGCAAAACCTTTTATTGATTACGGCTTCACGTC
 TTGCCAAAGAAAAGCACATTGATTGGCTTGGAAGCTGTGATTGAAGCTCATAAGGAGTTACCGGAACTAACCT
 TTGATATCTATGGTAGTGGTGGAGAAGATTCTGCTTACAGAGAAATTTGCAAAATCATCAGGCAGGACTATAT
 20 CCAACTCAAGGGGCATGCGGAACCTTTCGAGATTTATAGCCAGTATGAGGTCTACTTAACGGCTTCTACCAGCGA
 AGGATTTGGTCTGACCTTGATGGAAGCTATTGGTTTCAGGTCTACCTCTAATTGGTTTTGATGTGCTTATGGTAATC
 AGACCTTTATAGAGGATGGGCAAAATGGTTATTGATTCCAAGTTTCATCTGACCATGTAGAAAGACCAAAATCAAGC
 AAGCTTATGCCGCTAAGATTGTCAATTGTATCAAGAAAATCGTTTGAAGCTATGCGTGCTTCTTACCAAAAT
 TGCAGAAGGCTTCTTGACCAAGAAAATTTAGAAAAGTGAAGAAAACAGTAGAGGAGGTGCTCCATGATTGA
 25 MTIYNINLIGIWASSGVEYAQAYRAGVFRKLNLSKFIFDMLADNIQHLTANIGFDDNQVIWLYNHFTDKIAPTSVT
 VDDVLA YFGGEESHREKNGKVL RVFFDQDKFVTCYLVDENKDLVQHA EYVFKGNLRKDYFSYTRYCSEYFAPKDN
 VAVLYQRTFY NEDGTPVYDILMNQKKEEVYHFKDKIFYGKQAFVRAF MKSLNLNKS DVLDR ETGIGQVVFEEAQTA
 HLA VVHAEHYSENATNEDYILWNNYDYQFTNADKVDFFIVSTDRQNEVLQEQFAKYTOHQPKIVTIPVGSIDSLTDS
 30 SQGRKPFSLITASRLAKEKHIDWL VKAVIEAHKELPELTFDIYSGGSDSLREIHANHQAEDYIQLKGHAELSIIYSQYE
 VYLTASTSEGFLTLMEAGSGLPLIGFDPVYGNQTFIEDGQNGYLPSSSDHVEDQIKQAYAAKICQLYQENRLEAMRA
 YSYQLAEGFLTKEILEKWKTVEEVLHDZ

35

ID103 2292bp

40 ATGTCCTCTCTTTCCGGATCAAGAATTAGTAGCTAAAACAGTAGAGTTTCGTCAGCGTCTTTCCGAGGGAGAAAAGTC
 TAGACGATATTTTGGTTGAAGCTTTTGCTGTGGTGCCTGAAGCAGATAAGCGGATTTTAGGGATGTTTCTTATGA
 TGTTCAGTCATGGGAGCTATTGTCATGCACTATGGAAATGTTGCTGAGATGAATACGGGGGAAGGTAAGACCTT
 GACAGCTACCATGCCTGTCTATTTGAACGCTTTTTCAGGAGAAGGAGTGATGGTTGTGACTCTAATGAGTATTTA
 TCAAAGCGTGATGCCGAGGAAATGGGTCAAGTTTATCGTTTTCTAGGATTGACCATTGGTGTACCATTTACGGAAG
 45 ATCCAAAGAGGAGATGAAAGCTGAAGAAAAGAAAGCTTATCTATGCTTCGGATATCATCTACACAACCAATAGTA
 ATTTAGGTTTTGATTATCTAAATGATAACCTAGCCTCGAATGAAGAAGGTAAGTTTTTACGACCGTTTAACTATGT
 GATTATTGATGAAATGATGATATCTTGCTTGATAGTGACAAAACCTCTGATTATTGCGGGTTCTCCTCGTGTTC
 AGTCTAATTACTATGCGATCATTGATACACTTGTAAACAACCTTGGTGAAGGAGAGGATTATATCTTTAAAGAGGA
 GAAAGAGGAGGTTTGGCTCACTACTAAGGGGGCCAAGTCTGCTGAGAATTTCTAGGGATTGATAATTTATACAA
 50 GGAAGAGCATGCGTCTTTTGTCTCGTCAATTTGGTTTATGCGATTGAGCTCATAAGCTCTTTACTAAAGATAAAGGAC
 TATATCATTCGTGGAATGAGATGGTACTGGTTGATAAGGGAACAGGGCGTCTAATGGAATGACTAACTTCAA
 GGAGGTCTCCATCAGGCTATTGAAGCCAAGGAACATGTCAAATTTCTCTGAGACGCGGGCTATGGCCTCGATC
 ACCTATCAGAGTCTTTTAAAGATGTTTAAATAGATATCTGGTATGACAGGGACAGGTAAGGTGCGGAAAAAGAG
 TTTATTGAAACTTACAATATGTCTGTAGTACGCATTCCAACCAATCGTCCGAGACAACGGATTGACTATCCAGATA
 55 ATCTATATATCACTTTACCTGAAAAAGTGTATGCATCCTTGAAGTACATCAAGCAATACCATGCTAAGGGAAATCC
 TTTACTCGTTTTGTAGGCTCAGTTGAAATGTCTCAACTCTATTCTGCTCTCTTGTTCGTGAAGGGATTGCCATA
 ATGTCCTAAATGCTAATAATGCGGCGCGTGAGGCTCAGATTATCTCCGAGTCAGGTGAGATGGGGGCTGTGACAG
 TGGCTACCTCTATGGCAGGACGTGGTACGGATATCAAGCTTGGTAAAGGAGTCGACAGGCTTGGGGGCTTGATTG
 TTATTGGGACTGAGCGGATGGAAAGTCAGCGGATCGACCTACAAATTCGTGGCCGTTCTGGTCTGACGGGAGATC
 60 CTGGTATGAGTAAATTTTTGTATCCTTAGAGGATGATGTTTCAAGAAAATTTGGTCCATCTTGGGTGCATAAAAA
 GTACAAAGACTATCAGGTTCAAGATATGACTCAACCGGAAGTATTGAAAGGTCGTAATACCGGAACTAGTCGA
 AAAGGCTCAGCATGCCAGTGATAGTCTGGAAGTTTCAGCAGCTCGTCAGACTCTGGAGTATGCTGAAAGTATGAA
 TATACAAAGGATATAGTCTATAAAGAGAGAAATCGTCTAATAGATGGTTCTCGTGACTTAGAGGATGTTGTTGTG
 65 GATATCATTGAGAGATATACAGAAAGGAGGAGGCTGATCCTATGCTAGTCGTTGATGTTTACTTTTATTG
 TGACCAATATTAGTTTTCATGTTAAAGAGGTTCCAGATTATATAGATGTAACCTGACAAAACCTGCAAGTTCGTAGCTT
 TATGAAGCAGGTGATTGATAAAGAACTTTCTGAAAAGAAAGAAATTAATCAACATGACTTATATGAACAGTT

10

15

20

25

35

40

45

50

55

65

GLQAGFVQGDVVRPQMLKESDVISDLVPGYYPDDAVASRHQVASSQEHTYAHLLMEQGLKYLKSDGYAIFLAPSD
LLTSPQSDLLKEWLKEEASLVAMISLPENLFANAKQSKTIFILQKKNEIAVEPFVYPLASLQDASVLMKFENFQKWTQ
TEIZ

5 ID110 1902bp

ATGATTATTTTACAAGCTAATAAAATTGAACGTTCTTTTGCAGGAGAGGTTCTTTTCGATAATATCAACCTGCAGG
TTGATGAACGAGATCGGATTGCTCTTGTGGGAAAAATGGTGCAGGTAAGTCTACTCTTTTGAAGATTTTAGTTGG
AGAAGAGGAGCCAACTAGCGGAGAAATCAATAAGAAAAAGATATTTCTCTGTCTTACCTAGCCCAAGATAGCCG
10 TTTTGAGTCTGAAAAATACCATCTACGATGAAATGCTTCATGTCTTTAATGATTTCGCTCGGACGGAGAGACAACTG
CGTCAGATGGAGCTGGAGATGGGTGAAAAAGTCTGGTGAGGATTTGGATAAACTGATGTCAGATTATGACCGCTTA
TCTGAGAAATTTTCGCCAAGCAGGTGGCTTTACCTATGAAGCTGATATTCGAGCGATTTTGAATGGATTCAAGTTTG
ACGAGTCTATGTGGCAGATGAAAAATTGCTGAGCTTTCTGGTGGTCAAAATACTCGTTTGGCACTTGCCAAAAATGCT
15 CCTTGAAGGCCCAATCTCTTGGTCTTGGACGAGCCAACTAACCACTTGGATATTGAAACCACTCGCTGGCTAGA
GAATTACTTGGTAACTATAGCGGTGCCCTCATTATCGTCAGCCACGACCGTTATTTCTTGGACAAGGTTGCGACA
ATTACGCTAGATTGACCAAGCATTCCTTGGATCGCTATGTGGGGAATTAATCTCGTTTGTGCAATTGAAGGAGC
AAAAGCTAGTTACTGAGGCAAAAACTATGAAAAGCAACAGAAGGAAATCGCTGCTCTGGAAGACTTTGTCAATC
GCAATCTAGTTCTGCTTCAACGACTAAACGCTGCTCAATCTCGCCGTAAACAACTAGAAAAAATGGAGCGTTTGG
20 ACAAGCCTGAAGCTGGCAAGAAAGCAGCCAACTGACCTTCCAGTCTGAAAAAAGCTCGGGCAATGTTGTTTGA
CTGTTGAAAAATGCAGCTGTTGGCTATGACGGGGAAGTCTTGTCAACCTATCAACCTAGATCTTCGTAAGATGAA
TGGTGTCTGCTATCGTTGGTCCAAATGGTATCGGCAAGTCAACCTTTATCAAGTCTATTGTGGACCAAGATTCCTTTT
ATCAAGGGAGAAAAAGCGCTTGGCGCTAATGTTGAGGTTGGTTACTATGACCAAAACCAAGCAAGCTGACACCA
AGTAATACGGTGCTGGATGAACCTGGAATGATTCAAACCTGACACCAGAAGTTGAAATCCGCAACCGTCTTGGG
GCCTTCTCTTTTCTCAGGAGATGATGTTAAAAAATCAGTCGGCATGCTATCTGTTGGCGAAAAAGCTCGTTTGTCTT
25 TAGCTAAATTGTCTATGGAACAATAAATTTTGTATTCTGGATGAGCCGACCAACCACTTGGATATTGATAGTAA
GGAAGTGCTAGAAAAATGCCTTGATTGACTTTGATGGAACCTTGTCTGTTGTGCTAGTCATGATCGTTACTTTATCAAT
CGTGTGGCAACTCATGTTTGGAAATGTCTGAGAATGGTTCAACTCTTACCTTGGAGATTACGACTACTATGTTG
AGAAGAAAGCAACAGCAGAAATGAGTCAGACTGAGGAAGCTTCAACTAGCAATCAAGCAAGGAAGCAAGTCCA
GTCAATGACTATCAGGCCAGAAAGAAAGTCAAAAAAGGTTGCGAACTCATGCGACAAATCGAAAGTCTAGA
30 AGCTGAAATTGAAGAGCTAGAAAGTCAAGGCCAAGCCATTCTGAACAAATGTTGGAACAAACGATGCCGACA
AACTCATGGAATTACAGGCTGAGCTGGACAAATCAGCCATCGTCAGGAAGAAGCTATGCTTGAGTGGGAAGAAT
TATCAGAGCAGGTGTA

MIILQANKIERSFAGEVLFNDINLQVDERDRIALVGKNGAGKSTLLKILVGEEPTSGEINKKKDISLSYLAQDSRFESENT
15 IYDEMLHVFNDLRRTERQLRQMELEMGESGEDLDKLMSDYDRLSENFRQAGGFTYEADIRAILNGFKFDESMWQMK
IAELSGGQNTRLALAKMLLEKPNLLVLDEPTNHLDIETIAWLENYLVNYSALIIVSHDRYFLDKVATITLDTLKHSLDR
YVGNYSRFVELKEQKLVTEAKNYEKQKELAALEDVFNRLVRASTTKRAQSRKQLEKMERLDPKPEAGKKAANMTF
QSEKTSGNVVLTVENAAVGYDGEVLSPINLDRKMNVAIVGPNIGKSTFKISIVDQIPFIKGEKRFGANVEVGYDQ
40 TQSKLTPSNTVLDELWDFKLTPEVEIRNRLGAFVSGDDVKSVGMLSGGEKARLLAKLSMENNFLILDEPTNHL
DIDSKVELENALIDFDGTLFVSHDRYFINRVATHVLESENSTLYLGDYDYVEKKATAEMSQTTEASTSNQAKEAS
PVNDYQAQKESQKEVRKLMRQIESLEAEIELESQSQAISEQMLETNDADKLMELQAEKISHRQEEAMLEWEELSEQ
VZ

45 ID111 1179bp

ATGAATCGCTATGCAGTGCAGTTGATTAGCCGTGGGGCTATCAATAAAATGGGAAATATGCTCTATGATTATGGA
AATAGTGTCTGGTTGGCTTCTATGGGGACTATAGGACAGACAGTTTTAGGAATGTATCAGATTCTGAGCTCGTCA
CATCTATTCTCGTCAATCCCTTGGCGGAGTTATTTAGACCGTTTTTCTCGTCGTAAGATTTAATGACGGCAGAT
50 CTGTTTGTGGGATTCTTTGTCTGGCTATTTCTTTCATAAGGAATGATAGCTGGATGATTGGCGCTTTGATTGTTGC
TAACATTGTGCAAGGCTATTGCTTTTGCCTTTTCTCGCACAGCCAATAAAGCTATCATAACTGAAGTGGTGGAGAAA
GATGAGATTGTGATCTATAATCTCGCTTAGAGCTGGTTTGCAGGTTGTAGGTGTTAGCTCTCCTGTTCTTCTCTT
CCTTGTTTTACAGTTTGAAGTCTCCATATGACGCTACTGCTAGACTCGCTGACTTTTTTCATTGCTTTTGTCTAG
TGGCTTCTCTTCAAAAAGAGGAAGCAAAAGTTCAAGAGAAAAAGGCTTTTACTGGGAGAGATATTTTGTAGATA
TCAAGGATGGGTACACTATATCTGGCATCAGCAAGAAATTTTCTTCTTTTGTGCTGCTAGCTTCCAGCGTTAATTT
55 CTTTTTGCAGCTTTTGAATTTCTACTTCCCTTTTCAATCAGCTTTACGGGTGAGAGGAGCCTATGCAAGTATTT
TAACTATGGGGCTATTGGTTCCATCATTGGGGCTCTTCTAGCTAGTAAAAATTAAGCTAATATTTATAATCTTTT
GATTTTACTGGCTTTGACAGGTGTCGGAGTTTTATGATGGGATTACCACTTCCAACCTTTTCTTCTTTCTGGA
ATTTAGTTTGTGAATTGTTATGACGATTTTAAATTTCACTTTTACTCAAGTACAAACCAAGGTTGAGAGCGAA
TTTCTTGGAAAGAGTACTGAGTACAAATTTTACCTAGCTATTCTATTATGCCTATTGCAAAAGGATTTATGACAGT
60 CTTGCCAAGTGTCCATCTTTATCTTCTTGTATTATTGGACTTGGAGTTGTAGCCTTATATTCTTAGCTCTCGGAT
ATGTTCAAGCTCATTTGAAAAATTGATATA

MNRYAVQLISRGAINKMGNMLYDYGNSVWLASMGITGQTVLGMVQISELVTSLVNPFGGVISDRFSRRKILMTADLV
65 CGILCLAISFIRNDSWMIGALIVANTVQALAFASRTANKAITVEVKEDEIVIYNSRLELVQVVGVSPPVLSFLVLQFASL
HMTLLDSLTFIAFVLVAFLPKEEAKVQEKKAFTGRDIFVDIKDLHYIWHQOEIFFLLLVASSVNFFAAFELLPFSN

QLYGSEGA YASILT MGAIGSIH GALLASKIKANIY NLLILLALTGVGVFMMGLPLPTFLSFSGNLVCELFMTIFNIHFFTQV
QTKVESEFLGRVLSTIFTLAILFMPIAKGFMTVLPVSHLYSFLIIGLVVALYFLALGYVRTHFEKLIZ

ID113 2466bp

5 ATGCAAAATCAATTAATGAATTAACGAAAAATGCTGGAATTTTCCAGCAAAAACAAAAAATAAAAAATCA
GCTAGACCTGGCAAGAAAGGTTCAAGTACCAAAAAATCTAAAACTTAGATAAGTCAGCCATTTTCCAGCTATT
TTACTGAGTATAAAAGCCTTATTTAACITACTCTTTGTAAGTCTCGGTTTCTAGGAGGAATGTTGGGAGCTGGGATTG
10 CTTTGGGATACGGAGTGGCCTTATTTGACAAGGTTGGGTGCTCAGACAGAAGAAATGGTGAATCAGGTCAAGG
ACATCTCTTCTATTTTACAGAGATTACCTATTCGGACGGGACGGTGATTGCTTCCATAGAGAGTGATTGTTGCGCAC
TTCTATCTCATCTGAGCAAATTCGGAATCTGAAGAAGGCTATCATTGCGACAGAAGATGAACACTTTAAAGA
ACATAAGGGTGTAGTACCCAAGGCGGTGATTGCTGCGACCTTGGGGAATTTGTAGGTTTGGGTTCTCTAGTGGG
15 GGTCAACCTTGACCCAGCAACTAATTAACAGCAGGTGGTGGGGATGCGCCGACCTTGGCTCGTAAGGCGGCA
GAGATTGTTGGATGCTCTTGCCTTGAACGCGCCATGAATAAAGATGAGATTTTAAACGACCTATCTCAATGTGGCTC
CCTTGGCGGAAATAATAAGGGACAGAATATTGCAAGGGCTCGGCAAGCAGCTGAGGGAATTTTGGTGTAGATG
CCAGTCAGTTGACTGTTCTCAAGCAGCAATTTTAGCAGGACTTCCACAGAGTCCCATTACTCTCTCTATGA
AAACTACTGGGAGTTGAAGAGTGATGAAGACCTAGAAATTTGGCTTAAGACGGGCTAAGGCAGTTCTTACAGTAT
GTATCGTACAGGTGCATTAAAGCAAGACGAGTATTCTCAGTACAAGGATTATGACCTTAAACAGGACTTTTACC
20 ATCGGGCACGGTTACAGGAATTTACGAGACTATTTATACTTTACAACCTTGGCAGAAGCTCAAGAAGCTATGTAT
GACTATCTAGCTCAGAGAGACAATGTCTCCGCTAAGGAGTTGAAAAATGAGGCAACTCAGAAGTTTATCGAGAT
TTGGCAGCCAAGGAAATTGAAAAATGGTGGTTATAAGATTACTACTACCATAGATCAGAAAAATCTTCTGCCATG
CAAAGTGCGGTTGCTGATTATGGCTATCTTTAGACGATGGAACAGGTCGTGTAGAAGTAGGGAATGTCTTGATG
GATAACCAAACAGGTGCTATTTAGGCTTTGTAGGTGGTCGTAATTATCAAGAAAAATCAAATAATCATGCCTTTG
25 ATACCAAACGTTGCGCAGCTTCTACTACCAAGCCCTTGGCTACGGTATTGCTATTGACAGGGCTTGATGGG
AAGTGAACGATTCTATCTAACTATCCAACAACTTTGCTAATGGCAATCCGATTATGTATGCTAATAGCAAGGG
AACAGGAATGATGACCTTGGGAGAAGCTCTGAATATTCTGGAATATCCCTGCTTACTGGACCTATCGTATGCTC
CGTGAAGGGTGTGATGTCAAGGGTATATGGAAGATGGGTTACGAGATTCTGAGTACGGTATTGAGAGC
TTGCCAATGGGTGGTGGTATTGAAGTCACAGTTGCCAGCATACCAATGGCTATCAGACCTTAGCTAATAATGGA
30 GTTATCATCAGAACGATGTGATTCAAAGATTGAAGCAGCAGATGGTAGAGTGGTGTATGAGTATCAGGATAAA
CCGGTTCAAGTCTATTCAAAGCTACTGCGACGATTATGCAAGGATTGCTACGAGAAGTTTCTCTCTCGTGTGA
CAACAACCTTCAAGTCTAACCTGACTTCTTTAAATCCTACTCTGGCTAATGCAGATTGGATGGGAAGACTGGTAC
AACCAACCAAGACGAAAAATATGTGGCTCATGCTTTGACACCTAGATTAAACCTAGGTGGCTGGATTGGGCATGA
TGATAATCATTCTATTGTACGTAGAGCAGGTTATTCTAATACTCTAATTACATGGCTCATCTGGTAAATGCGATT
35 CAGCAAGCTTCCCAAGCATTGGGGGAACGAGCGCTTTGCTTTAGATCCTAGTGTAGTGAAATCGGAAGTCTTG
TAGGAAACAGGTCAAAAAACAGAGAAGGTTCTGTTGAAGGAAAAAGTAGAGGTCACAGGTTCCGACTGTTACC
AGCTATTGGGCTAATAAGTCAGGAGCGCCAGCGACAAGTTATCGCTTTGCTATTGGCGGAAGTGATGCGGATTAT
CAGAATGCTTGGTCTAGTATTGTGGGGAGTCTACCACTCCATCCAGCTCCAGCAGTTCAAGTAGTATTCTAGCG
40 ATAGCAGTAACCTCAAGTACTACACGACCTTCTTCTCAAGGGCGAGACGATAA

40 MQNQLNELKRKMLEFFQKQKNKKSARPKKGSSTKSKTLDKSAIFPAILLSIKALFNLLFVLGFLGGMGLGAGIALGY
GVALFDKVRVPQTEELVNQVKDISSIEITYSDGTVIASIESDLLRTSISSEIQUENLKKAILATEDEHFKEHKGVVPKAVIR
ATLGKFVGLGSSSGSTLTQQLIKQVVGDAPTLARKAAEIVDALALERAMNKDEILTYLVNAPFGRNKNKGQNIAGA
RQAAEGIFGVDASQLTVPQAAFLAGLPSPITSPYENTGELKSDLEDLEIGLRRAKAVLYSMYRTGALSKEYSQYKDY
45 DLKQDFLPSGTVTGISRDYLYFTTLAEAQERMYDYLAQRDNVSAKELKNEATQKFYRDLAAKEENGKYKITTTIDQKI
HSAMQSAVADYGYLLDDGTGRVEGVNMDNQTGAILGFVGGRNQENQNNHAFDTRSPASTTKPLLYAGLAIDQG
LMGSETILSNYPNPFANGNPIMYANSKGTGMMTLGEALNYSWNIPAYWYRMLREKGVDDVKGYMEKMGYIEPEYGIE
SLPMGGGIEVTVAQHTNGYQTLANNGVYHQKHVISKIEADGRVVEYQDKPVQVYSKATATIMQGLLREVLSSRVTT
TFKSNLTSLNPTLANADWIGKTGTTNQDENMWMMLSTPRLTLGGWIGHDDNHSLSRRAGYSNNSNYMAHLVNAIQQA
50 SPSIWGNERFALDPSVVKSEVLKSTGQKPEKVSVEGKEVEVTGSTVTSYWANKSGAPATSYRFAIGGSDADYQNAWSSI
VGS LPTPSSSSSSSSSSSSSSSSSSSTTRPSSSRARRZ

ID114 1974bp

55 ATGAAAAAATTTTATGTAAGTCCAATTTTCTATTCTAGTAGGATTGATTGCGTTTGGAGTCTTATCCACTTTTAT
TATTTTGTATAATAATCTGTTGACGGTTTAAATTTGTTTCTTTTGTAGGAGGCTATGTTTTTATTTAAGAA
ACTGAGAGTGCAATTATACAAGGAGTGATGTAGAACAGATACAGTATGTAAACCAACCAAGCGGAAGAAAGTTTGAC
AGCTCTATTGGAACAGATGCCTGTAGGTGTTATGAAATTGAATTTATCTTCTGGAGAGGTTGAGTGGTTTAAATCCC
60 TATGCTGAATGATTTTGACCAAGGAAGATGGTGATTGATTAGAAAGCTGTTCAAACGATTATCAAGGCTTCAG
TAGGAAATCCGCTACTTATGCAAGCTTGGTGAGAAGCGTTATGCTGTTTATGATGGATGCTTCTCCGGTGTGTTT
GTATTTGTAGATGTATCCAGGGAACAAGCCATAACAGATGAATTGGTAACAAGTAGACCAGTGATTGGGATTGT
CTCTGTGGATAATTATGATGATTTGGAGGATGAAACTTCTGAGTCAGATATTAGTCAAAATCAATAGTTTGTAGCT
AATTTTATCAGAGTTTTCAGAAAAACACATGATGTTTCTCGTCGGGTAAGTATGGATCGATTATCTATTATAC
TGACTACACGGTGCTTGAGGGCTTGATGAATGATAAATTTCTGTTATTGATGCTTTCAGAGAAGAGTCGAAACAG
65 AGACAGTTGCCCTTGACCTTAAGTATGGGATTTCTATGGCGATGGAATCATGATGAGATAGGGAAGGTTGCTT
TGCTCAATTTGAACCTTGGCTGAAGTACGTGGTGGCGACCAAGGTGGTGTAAAGGAAAACGACGAAACGAAAAATC

5 CAGTTTATTTTGGTGGTGGGTCTGCTGCTTCAATCAAGCGTACACGGACTCGTACGCGCGCTATGATGACAGCTAT
TTCAGATAAGATTCCGAGTGTAGATCAGGTTTTGTAGTCGGTCACAAAAATTTAGACATGGATGCTTTGGGCTCT
GCTGTAGGTATGCAGTTGTTCCGCAGCAATGTGATTGAAAAATAGCTATGCTCTTTATGATGAAGAACAAATGTCTC
CAGATATTGAACGAGCTGTTTCATTCATAGAAAAAGAGGATTACGAAGTTGTTGCTGTTAAGGATGCAATGG
GGATGGTGACCAATCGTCTTTGTTGATTCTGTAGACCATTCAAAGACAGCCTTAACATTATCAAAAGAAATTTTA
TGATTTATTTACCCAAACCATTGTTATTGACCACCATAGAAGGGATCAGGATTTTCCAGATAATGCGGTTATTACT
TATATCGAAAGTGGTGCAAGTAGTGCCAGTGAGTTGGTAACGGAATTGATTCACTTCCAGAAATTTCTAAGAAAAAT
CGTTTGAGTCGTATGCAAGCAAGTGTCTTGATGGCTGGTATGATGTTGGATACTAAAAATTTCACTCGCGAGTAA
10 CTAGTCGGACATTTGATGTTGCTAGCTATCTCAGAACGCGCGGAAGTGATAGTATTGCTATCCAGGAAATCGCTGC
GACAGATTTTGAAGAAATATCGTGAGGTCAATGAACITATTTTACAGGGGCGTAAATAGGTTGAGATGTAATA
GCAGAGGCTAAGGACATGAAATGCTATGATACAGTTGTTATTAGTAAGGCAGCAGATGCCATGTTAGCCATGTCA
UGTATTGAAGCGAGTTTGTCTTGGCAAGAATACACAAGGATTTATCTCTATCTCAGCTCGAAGTCGTAGTAAAC
TGAATGTACAACGAGATTATGGAAGAGTTAGGCGGTGGAGGCCACTTTAATTTGGCAGCAGCTCAAATTAAGATG
15 TAACCTTGTGAGAAGCAGGTGAAAACTGACAGAAATGTATTAAATGAAATGAAGGAAAAGGAGAAAAGAA
TGA

20 MKKFYVSPFIPILVGLIAGVLTSTFIIVNNLLTVLILFLVGGYVFLFKLRVHYTRSDVEQIYVNHQAEESLTALLE
QMPVGVMLNLSSGEVEWFPYAEILITKEDGDFLEAVQTIKASVGNPSTYAKLGEKRYAVHMDASSGVLYFVDVS
REQAITDELVTSRPVIGIVSVNDYDDLEDETSDESISQINSFVANFISEFSEKHMMSRRVSMDFYLFYDVTLEGLMN
DKFSVIDAFREESKQRQLPLTSMGFSYGDGNHDEIGKVALLNLNLAEVRRGGDQVVVKENDETKNPVYFGGSAASIK
RTRTRTRAMMTAISDKIRSVQVFFVGHKNLMDALGSAVGMQLFASNVNIENSYALYDEEQMSPDIERAVSFIEKEGV
TKLLSVKDAMGMVNTNRSLLILVDHSKTALTLSKEFYDLFTQTTVIDHRRDQDFPDNAVITYIESGASSASELVTELIQFQ
NSKKNRLSRMQASVLMAGMMLDTKNFTSRVTSRTFDVASYLRTGRSDSIAIQEIAATDFEYREVNELILQGRKLGSDV
25 LIAEAKDMKCYDTVVISKAADAMLAMSGIEASFVLAKNTQGFISISARSRLNVQRIMEELGGGGHFNLAQAQKDV
LSEAGEKLTEIVL NEMKEKEKEEZ

ID115 663bp

30 ATGAAGTCTTGTATGTGGGCAGACTATGAAGACTGTTTTAACTTTTAGTAGTCTCTTACTTCTGAGGAATGATG
ACTCTTGTCTTTGTTCAGACTGTGATTCTACTTTTGAAGAATTGGGGAAGAGAACTGTCCAAATTTGATGAAAA
AGAGTTGTCAACAAAGTGTCAAGATTGTCACTTTGGTGTAAAGAGGGAGTTGAAGTCAGTCATAGAGCGGATTTT
TACTTACAATCAAGCTATGAAGGATTTTTTCAGTCGGTATAAGTTTGAAGGAGACTTCTGTTAAGAAAAAGTTTTC
GCTTCATTTTTAAGTGAGGAGTTGAAAAAGTACAAAGAGTATCAATTTGTTGTAATCCCTTAAGTCTCGATAGAT
35 ATGCTAATAGAGGATTTAATCAGGTTGAGGGCTTGGTAGAGGCAGCAGGCTTGAAGTATCTGGATTTATTAGAGA
AAAGAGAAGAGAGCCAGTTCTTCTAAAAATCGTTCAGAGCGCTTGGGACAGAACTTCTTCTTTATTTAAAA
GTGGAGTCACTATTCTAAAAAAATCTACTTATAGATGATATCTATACTACAGGAGCAACTATAAATCGTGTAA
GAACTGTTGGAAGAAGCTGGTGCTAAGGATGTAACAACTTTCCCTTGAAGATGA

40 MKCLLCGQTMKTVLTFSSLLLLRNDSDCLDCDSTFERIGEENCPCNMKTELSTKCQDCQLWCKEGVEVSHRAIFTY
NQAMKDFFSRYKFDGDFLLRKVFASFLSEELKKYKEYQFVVIPLSPDRYANRGFNQVEGLVEAAGFEYLDLLEKREER
ASSSKNRSERLGTLPFFIKSGVTIPKKILLIDDIYTTGATINRVKKLLEEAGAKDVKTFSLVZ

ID116 1299bp

45 ATGAAAGTAAATTTAGATTATCTCGGTGTTTTATTTACTGAGAATGAATTAACAGAAGAAGAACGTCAGTTGGCG
GAGAAACTCCAGCAATGAGAAAGGAGAAAGGGGAACTTTCTGTCAACGCTGTAATAGTACTATTCTAGAAAGAA
TGGTATTTGCCCATCGGTGCTTACTATTGTGAGAGTGCTTGTGATGAAGCGAGTCAGAAGTGATCAAACTTTAT
ACTATTTCCGCAGGAGGATTTTCCAAAGCAAGATGTTCTCAATGGCGCGGCCAATTAACCTCTTTTCAAGAGAA
50 GGTGTGAGAGGGATTGCTTCAAGTAGTAGACAAGCAAGCAACCTTAGTTCATGCGGTAAACAGGAGCTGGAAA
GACAGAAATGATTTATCAAGTAGTGCTAAAGTGATCAATGCGGGTGGTGCAGTGTGTTGGCTAGTCTCGCAT
AGATGTTTGTGTTGAGCTGTACAAGCGCCTGCAACAGGATTTTCTTGGCGGATAGCTTTGCTACATGGAGAATCG
GAACCTTATTTTGAACACCACTAGTTGTTGCAACAACCCATCAGTTATTGAAGTTTTATCAAGCTTTTGAITTGCT
GATAGTGGATGAAGTAGATGCTTTTCTTATGTTGATAATCCCATGCTTTACCACGCTGTCAAGAATAGTGTAAG
55 GAGAATGGATTGAGAATCTTTTAAACAGCGACTTCGACCAATGAGTTAGATAAAAAGGTCCGTTTAGGAGAACTA
AAAAGACTGAATTTACCGAGACGGTTTCATGGAATCCGTTGATTATCCAAAACCAATTTGGTTATCGGATTTTA
ATCGCTACTTAGACAAGAATCGTTTGTCAACAAAGTTAAAGTCTATATTGAGAAGCAGAGAAAGACAGCTTATC
CGTTACTCATTTTTGCTTCAGAAATTAAGAAAGGGGAGCAGTTAGCAGAAATCTTACAGGAGCAATTTCCAAATG
AGAAAAATGGCTTTGTATCTTCTGTAAACAGAGGATCGATTAGAGCAAGTACAAGCTTTTTCGAGATGGAGAACTGA
60 CAATACTATCAGTACGACAATCTTGGAGCGCGGAGTTACCTTCCCTTGTGTGGATGTTTTCGTAGTAGAGGCCAA
TCATCGTTTGTGTTACCAAGTCTAGTTTGAATTCAGATTGGTGGACGAGTTGGACGAAGCATGGATAGCCGACAGGA
GATTTGCTTTTCTTCATGATGGGTTAAATGCTTCAATCAAGAAGGCGATTAAGGAAATTCAGATGATGAATAAGG
AGGCTGGTCTATGA

65 MKVNLDYLGRLFTENELTEERQLAEKLPAMRKEKGKLFQRCNSTILEEWYLPIGAYYCRECLLMKRVRSQDTLYYF
PQEDFPKQDVLKWRGQLTPFQEKVSEGLLQVVDKQKPTLVHAVTGAGKTEMITYQVAVKVINAGGAVCLASPRIDVCL

ELYKRLQQDFSCGIALHGESEPYFRTPLVVATTHQLLKFYQAFDLLIVDEVDAFPYVDNPMLYHAVKNSVKENGLRIF
LTATSTNELDKKVRLEGELKRLNLPFRFHGNPLIPKPIWLSDFNRYLDKNRLSPKLKSYIEKQKRTAYPLLFASEKKGE
QLAEILQEOPNEKIGFVSSVTEDRLEQVQAFRDGELITLISTILERGVTFPCVDVFVVEANHRLFTKSSLIQIGGRVGRS
MDRPTGDLFFHDGLNASIKKAIKEIQMMNKEAGLZ

5

ID117 870bp

ATGCAAAATCAAAAAAGTTTTAAGGGGCAGTCTCCCTATGGCAAGCTGTATCTAGTGGCAACGCCGATTGGCAAT
CTAGATGATATGACTTTTCGTGCTATCCAGACCTTGAAAGAAGTGGACTGGATTGCTGCTGAGGATACGCCGAAT
ACAGGGCTTTTGCTCAAGCATTGACATTTCCACCAAGCAGATCAGTTTTTCATGAGCACAAATGCCAAGGAAAAA
ATTCTGATTGATTGGTTTCTTGAAAGCAGGGCAAAGTATTGCTCAGGTCTCTGATGCCGTTTTCCTAGCATTT
CAGACCTGGTCATGATTTAGTTAAGGCAGCTATTGAGGAAGAAATTCAGTTGTGACAGTTCCAGGTGCTCTGTC
AGGAATTTCTGCCTTGATTGCCAGTGGTTAGCGCCACAGCCACATATCTTTACGGTTTTTACCGAGAAAAATCA
GGTCAGCAGAAGCAATTTTTGGCTTGAAAAAGA'ITATCTGAAACACAGATTTTTATGAATCACCTCATCGTG
TAGCAGACACGTTGGAAAAATATGTTAGAAGTCTACGGTGACCGCTCCGTTGCTTGGTCAGGGAATTGACAAAA
TCTATGAAGAATACCAACGAGGTACTATCTCTGAGTATTAGAAAGCATTGTGTAAGCGCCACTCAAGGGCGAAT
GTCTTCTCATTGTTGAGGGTGCCAGTCAGGGTGTGGAGGAAAGGACGAGGAAGACTTGTTCGTAGAAATTCAAA
CCCGCATCCAGCAAGGTGTGAAGAAAAACCAAGCTATCAAGGAAGTCGTAAGATTACCAGTGAATAAAAGTC
AGCTCTACGCTGCCTACCAGACTGGGAAGAAAAACAATAA

10

15

20

MOIQKSFQSPYKLYLVATPIGNLDDMTFRAIQTLEKVDWIAAEDTRNTGLLLKHFDISTKQISFHEHNAKEKIPDLI
GFLKAGQSIAQVSDAGLPSISDPCHDLVKAJEEELAVVTPGASAGISALIASGLAPQPHIFYGLPRKSGQKQOFFGLKK
DYPETQIFYESPHRVADTLENMLEVYGDERSVVLVRELTKIYEYQRTISELLESIAETPLKGECLLIVEGASQGVEEKDE
EDLFVEIQTRIQQGVKKNQAIKEVAKIYQWNKSLYAAHYDWEKQZ

25

ID118 345bp

ATGATAAAGAAAGGAAAGGGCTGTTTTATGGACAAAAAGAATTATTTGACGCGCTGGATGATTTTTCCCAACAA
TTATTGGTAACCTTAGCCGATGTGGAAGCCATCAAGAAAAATCTCAAGAGCCTGGTAGAGGAAAAATACAGCTCTT
CGCTTGGAAAAATAGTAAGTTGCGAGAACGCTTGGGTGAGGTGGAAGCAGATGCTCCTGTCAAGGCCAAGCATGTT
CGCGAAAGTGTCCGTCGTATTTACCGTGATGGATTTACCGTATGTAATGATTTTTATGGACAACGTCGAGAGCAGG
ACGAAGAATGTATGTTTTGTGACGAGTTGTTATACAGGGAGTAA

30

35

MIKKKGKCFMDKKELFDALDDFSQQLLVTLADVEAIKKNLKSLEVENTALRLENSKLRERLGEVEADAPVKAKHVRES
VRRIYRDGFHVCNDFYQRRREQDEECMFCELLYREZ

ID119 639bp

ATGTCAAAAGGATTTTTAGTCTCTCTTGAGGGACCAGAGGGAGCAGGCAAGACCAGTGTTTTAGAGGCTCTGCTA
CCAATTTTAGAGGAAAAAGGAGTAGAGGTGTTGACGACCCGTGAACCTGGCGGAGTCTTGATTGGGGAGAAGATT
CGGGAAGTGATTTTGGATCCAAGTCATACTCAGATGGATGCTAAAACAGAGCTACTTCTCTATATTGCCAGTCGCA
GACAGCATTGGTGGAAGAAAGTTCTCCAGCCCTTGAAGCTGGCAAGTTGGTCATCATGGATCGTTTTATCGATAG
TTCTGTGCTATCAGGGATTTGGTCTGGCTTAGATATTGAAGCCATTGACTGGCTCAATCAGTTTGGCAGAGAT
GGCCTCAAACCCGATTTGACACTCTATTTGACATCGAGGTGGAAGAAGGGCTGGCTCGTATTGCTGCTAATAGTG
ACCGCGAGGTTAATCGTTTGGATTGGAAGGGTTGGACTTGCATAAAAAAGTTTCGTAAGGCTACCTTTCTCTCT
GGATAAAGAGGGAAATCGCATTGTCAAGATTGATGCTAGTCTCCCTTTGGAGCAAGTTGTGGAACTACCAAGGC
TGCTCTGTTTGACGGAATGGGCTTGCCAAATGA

40

45

50

MSKGFLVSLEGPEGAGKTSVLEALLPILEEKGVLEVTREPGGVLIGEKIREVILDPSTQMDAKTELLYIASRRQHLVE
KVLPALEAGKLVIDRFDSSVAYQGFGRGLDIEAIDWLNQFATDGLKPDLTLYFDIEVEEGLARIAANSREVNRLDL
EGLDLHKVKVRQGYLSLLDKEGNRIVKIDASLPLEQVVETTKAVLFDGMGLAKZ

ID120 408bp

ATGGTAGAACAAAGAAAAATCAATTACCATGAAAGATGTTGCTTTAGAAGCAGGAGTTAGTGTGGAACTGTTTCA
CGTGTAATTAATAAAGAAAAAGGCATTAAGAAAGTAACCTTGAAAAAAGTGGAAACAAGCGATTAAAACTTTGAAT
TACATTCCAGATTACTACGCTAGAGGAATGAAAAAAATCGAACAGAAACGATTGCAATCATTTGACCAAGTATC
TGGCATCCCTTCTTTTCAAAATTTGCTATGCTATGTTGAAAAATGAAGTCTATAAGAGAAATAACAAATTACTCTTAT
GTTCTATCAATGGTACAAATAGAGAGCAAGACTATCTGGAGATGTTGCGTCATAATAAAGTTGATGGAGTGGTTG
CCATTACCTATAGGCCAATTGAACATTACTTGACGTGAGGAATCCCTTTGTTAGTATTGACCGCACATACTCAGA
GATTGCCATTCTTGTGTTTCA

60

65

MVEQRKSITMKDVALEAGVSVGTVSRVINKEKGIVELTKKVEQAIKTLNYIPDYARGMKKNRTETIAIIVPSIWHPPF
SEFAMHVENEVYKRNNKLLCSINGTNREQDYLEMLRHNKVDGVVAITYRPIEHLTSGIPFVSIDRTYSEIAPCVS

ID121 285bp

5 ATGAATATATTTAGAACAAAGAATGTTAGTTTAGATAAAAACAGAGATGCATAGGCATTTGAAGTTATGGGATTTG
ATTTTGTGGGTATCGGAGCCATGGTAGGGACAGGCGCTTTACAATCACAGGTAAGTGCAGCTGCAACACTTGGCTG
GCCCAGCCCTAGTGATTTCAATCGTTATTTCTGCTTGTGTGGGATTATCAGCCCTCTTTTTGCAGAATTTGCC
TCGGAGTACCCGCTACAGGAGGTGCCTATAGTTACCTCTATGCTATCTTAGGAGAATTCCTGCGCTGGTGGCTG
GTTGGTTAACCATGATGGAGTTCATGACAGCCATATCAGGCGTAGCTTCGGGTTGGGCAGCTTATTTTAA

10 MNIFRTKNVSLDKTEMRHLKLWDLILLGIGAMVGTGVFTTGTAAATLAGPALVISIVISALCVGLSALFFAEFASRPV
ATGGAYSYLAILGEFPWLAGWLTMMEFMTAISGVASGWAAAYF

ID124 1311bp

15 ATGAAATCAAGAGTAAAGGAAACGAGTATGGATAAAATTGTGGTTCAAGGTGGCGATAATCGTCTGGTAGGAAGC
GTGACGATCGAGGGAGCAAAAAATGCAGTCTTACCCTTGTGGCAGCGACTATTCTAGCAAGTGAAGGAAAGACC
GTCTTGCAGAAATGTTCCGATTTTGTGCGATGTCTTTATTATGAATCAGGTAGTTGGTGGTTTGAATGCCAAGGTTG
ACTTTGATGAGGAAGCTCATCTTGTCAAGGTGGATGCTACTGGCGACATCACTGAGGAAGCCCCCTTACAAGTATG
TCAGCAAGATGCGCGCTCCATCGTTGTATTAGGGCCAATCCTTGCCCGTGTGGGTGATGCCAAGGTATCCATGCC
20 AGGTGGTTGTACGATTGGTAGCCGTCTATTGATCTTCATTTGAAAGGTCTGGAAGCTATGGGGTTAAGATTAGT
CAGACAGCTGGTTACATCGAAGCCAAGGCAGAACGCTTGCATGGTGCTCATATCTATATGGACTTCCAAAGTGTG
GTGCAACGCAGAACTTGATGATGGCAGCGACTCTGGCTGATGGGTGACAGTGATTGAGAATGCTGCGCGTGAGC
CTGAGATTGTTGACTTAGCCATTCTCCTTAATGAAATGGGAGCCAAGGTCAAAGGTGCTGGTACAGAGACTATAA
CCATTACTGGTGTGAGAACTTCATGGTACGACTCACAATGTAGTCCAAGACCGTATCGAAGCAGGAACCTTTAT
GGTAGCTGCTGCCATGACTGGTGGTGTGCTTGTATTGAGACGCTGTCTGGGAGCAACCCGCTCCCTTGTATTGCC
25 AAGTTACTTGAAATGGGTGTTGAAGTAATTGAAGAAGACGAAAGGAATTCGTGTTCTCAACTAGAAAACTCA
AAAGCTGTTTCAATGTGAAACCTTGCCCAACCCAGGATTTCCAACAGATATGCAGGCTCAATTTACAGCCTTGATGA
CAGTTGCAAAAGGCGAATCAACCATGGTGGAGACAGTTTTCGAAAATCGTTTCCAAACCTAGAAGAGATGCGCCG
CATGGGCTTGCACTTCTGAGATTATCCGTGATACAGCTCGTATTGTTGGTGGACAGCCTTTCAGGGAGCAGAAAGT
30 CTTTCAACTGACCTTCGTGCCAGTGCAGGCTTGATTTTGACAGGTTTGGTAGCACAGGAGAACTGTGGTCGTA
AATTGGTTCACTTGGATAGAGGTTACTACGGTTTCCATGAGAAGTTGGCGCAGCTAGGTGCTAAGATTACAGCGAT
TGAGGCAAGTGATGAAGATGAATAA

35 MKSRVKETSMKIVVQGGDNRLVGSVTIEGAKNAVLPLAATLASEGKTVLQNPILSDVFIMNQVVGGLNAKVDFD
EEAHLVKVDATGDTTEEAPYKYVSKMRASIVVLGPILARVGHAKVSMPGGCTIGSRPIDLHLKGLEAMGVKISQTAGYIE
AKAERLHGAIHYMDFPSVGATQNLMMAAATLADGVTVIENAAAREPEIVDLAILLNEMGAKVKGAGTETTTITGVEKLHG
TTHNVVQDRIEAGTFMVAAMTGGDVLIRDAVWEHNRPLIAKLLLEMGVEVIEDEGIRVRSQLENLKAHVHKTLPHP
GFPTDMQAQFTALMTVAKGESTMVETVFENRFQHLEEMRRLHSEIURDTARJVGQPLQGAEVLSTDLRASAALLIL
TGLVAQGETVVGKLVHLDRGYGFHEKLAQLGAKIQRIEASDEDEZ

ID125 1101bp

40 ATGTTATTAGCGTCAACAGTAGCCTTGTCAATTTGCCCAAGTATTGGCAACTCAAGCAGAAGAAGTCTTTGGACTG
CACGTAGTGTGAGCAAAATCCAAAACGATTGACTAAAACGGACAACAAAAAAGTTATACCGTACAGTATGGTG
45 ATACTTTGAGCACCATTGCAGAAGCCTTGGGTGTAGATGTACAGTGCTTGCGAATCTGAACAAAATCACTAATAT
GGACTTGATTTTCCAGAACTGTTTGTGACAACGACTGTCAATGAAGCAGAAGAAGTAACAGAAGTTGAAATCCA
AACACCTCAAGCAGACTCTAGTGAAGAAGTGACAACTGCGACAGCAGATTTGACCACTAATCAAGTGACCTTGA
TGATCAAACTGTTACGGTTGCAGACCTTTCTCAACCAATTGCAGAAGTTACAAAGACAGTGATTGCTTCTGAAGAA
GTGGCACCATCTACGGGCATTCTGTCCCAGAGGAGCAAAACGACCGAAACAACCTCGCCAGTTGCAGAAGAAGCT
50 CCTCAGGAAACGACTCCAGCTGAGAAGCAGGAAACACAAACAAGCCCTCAAGCTGCATCAGCAGTGAAGCAAC
TACAACAAGTTCAGAAGCAAAAGAAGTAGCATCATCAATGGAGCTACAGCAGCAGTTTCTACTTATCAACCAGA
AGAAACGAAAGTAATTTCAACAACCTACGAGGCTCCAGCTGCGCCCGATTATGCTGGACTTGCAAGTACAAAATC
TGAAAATGCAGGTCTTCAACCACAAACAGCTGCCTTTAAWGAAGAAATTGCTAAGTTGTTTGGCATTACATCCTTT
AGTGGTTATCGTCCAGGAGACAGTGGAGATCACGGAAGGTTTGGCTATCGACTTTATGGTACCAGAACGTTCA
55 GAATTAGGGGATAAGATTGCGGAATATGCTATTCAAAATATGGCCAGCCGTGGCATTAGTTACATCATCTGGAAA
CAACGTTTCTATGCTCCATTGATAGCAAAATATGGGCCAGCTAACACTTGGAAACCAATGCCAGACCGTGGTAGT
GTGACAGAAAATCACTATGATCACGTTTCAATGAATGGATAA

60 MLLASTVALSFAPVLATQAEVLTARSVEIQNDLTKTDNKTSTYTVQYGDTLSTIAELGVDVTVLANLNKITTMDL
IFPETVLTTTVNEAEVTEVEIQTPQADSSEVTTATADLTNNQVTVDDQTVQVADLSQPIAEVTKTVIASEEVPSTGTS
VPPEQTTETRPVAAEAPQETTPAEKQETQTSQAASAVEATTTSSAEKEVASSNGATAAVSTYQPEETKVISTYGEAPA
APDYAGLAVAKSENAGLQPTAAFKKLLTCLALHPLVIVQETVEITEKVWLSLWYQNVQNZGIRLRNMLFKIWP
65 VALVTSSGNNVSMHLSIANMGQLTLGTQCQTVVZQKITMITFTFQZMD

ID126 1281bp

5 TTGTTTAAGAAAAATAAAGACATTCTTAATATTGCATTGCCAGCTATGGGTGAAAACCTTTTTCAGATGCTAATGG
GAATGGTGGACAGTTATTTGGTTGCTCATTTAGGATTGATAGCTATTTCCAGGGGTTTCAGTAGCTGGTAATATTAT
CACCATTTATCAGGCGATTTTCATCGCTCTGGGAGCTGCTATTTCCAGTGTTATTTCAAAAAGCATAGGCGAGAAA
10 GACCAGTCGAAGTTGGCCTATCATGTGACTGAGGCGTTGAAGATTACCTTACTATTAAGTTTCCTTTTAGGATTTT
TGTCATCTTCGCTGGGAAAGAGATGATAGGACTTTTGGGGACGGAGAGGGATGTAGCTGAGAGTGGTGGACTGT
ATCTATCTTTGGTAGGCGGATCGATTGTTCTTAGGTTAATGACTAGTCTAGGAGCCTTGATTCTGTCAACGCA
TAATCCACGCTCTGCCTCTCTATGTTAGTTTTTATCCAATGCCTTGAATATTTCTTTTCAAGTCTAGCTATTTTGT
TCTGGATATGGGGATAGCTGGTGTGCTTGGGGGACAATTGTGTCTCGTTGGTTGGTCTTGTGATTTTGTGGTCAC
AATTAACCTGCCTTATGGGAAGCCAACTTTTGGTTAGATAAGGAACTGTTGACCTTGGCTTTACCAGCAGCTGG
15 AGAGCGACTTATGATGAGGGCTGGAGATGTAAGTATGCTTGCCTTGGTCTGTTCTTTGGGACGGAGGCACTGCT
GGGAATGCAATCGGAGAAGTCTTGACCGATTAACTATATGCCTGCCTTGGCGTCTGCTACGGCAACGGTCATG
CTGTTGGCCCGAGCAGTTGGAGAGGATGATTGGAAGAGATTGCTAGTTTGAGTAAACAAACCTTTTGGCTTCTC
TGTTCTCATGTTGCCCTGTCTTTAGTATATATGCTTGGGTGTACCTAACTCATCTCTATACGACTGATTCT
CTAGCGGTGGAGGCTAGTGTCTAGTGACACTGTTTCACTACTTGGGACCCCTATGACGACAGGAACAGTCATCT
20 ATACGGCAGTCTGGCAGGGATTAGGAAATGCACGCCCTCCCTTTTATGCGACAAGTATAGGAATGTGGTGTATCC
GCATTGGGACAGGATATCTGATGGGGATTGTGCTTGGTTGGGGCTTGCCTGGTATTTGGGCAGGGTCTCTTGGGA
TAATGGTTTTCGCTGGTTATTTCTACGCTATCGTTACCAGCGCTATATGAGCTTGAAGGATAG

LFKKNKDILNIALPAMGENFLQMLMGMVDSYLVHLGLIAISGVSVAGNITTYQAIFFALGAAISSVISKSIGQKDQSKLA
YHVTEALKITLLLSFLLGFLSIFAGKEMIGLLGTERDVAESGGLYLSLVGGSIIVLLGLMTSLGALIRATHNPRPLPVYSFL
25 SNALNIFSSLAIFVLDMDIAGVAVGTTIVSRVLVGLVILWSQLKLPYGKPTFGLDKELLTLALPAAGERLMMRAGDVIIL
LVVSFGTEAVAGNAIGEVLTQFNYPAGFVATATVMLLARAVGEDDWKRVASLSKQTFWLSFLMLPLSFSIYLVGP
LTHLYTDSLAVEASVLTFLSLLGTPMTTGTVIYAVWQGLGNARLPFYATSIGMWCIRGTGYLMGIVLGWGLPGIWI
AGSLDNGFRWFLRYRYQRYMSLKGZ

ID127 894bp

GTGGGAAGAATTATCAGAGCAGGTGTAAAGATGGAACATCTTGGAAAAGTATTTCTGTAATTTTGAACAAGTGGA
AATTATCTTTAAAGGAAGCAGCAGGCGAATCCTGCTCTACCTCTCAGTTATCTCGCTTTGAGCTTGGGGAGTCTG
35 ACCTGGCAGTCTCCCGTTTCTTTGAGATTTTGGATAACATTCATGTAACAATCGAAAAATTCATGGATAAGGCAAG
GAATTTTCATAATCATGAACATGTGTCTATGATGGCACAGATTATCCCACTTTACTATTCAAACGATATTGCAAGG
TTTCAAAGCTTCAAAGAGAACAACCTTGAAGAGTCTAAGAGTTTCGACGACTCCCTTTATTTTGAAGTGAAGTGA
TTTTGCTACAAGGTCTGATTTGTCAAAGAGATGCGAGTTATGATATGAAGCAGGATGATTGGGTAAGGTAGCAG
ATTATCTCTTCAAACAGAGAAGTGGACCATGTATGAGTTGATTCTTTTCGGTAACCTCTATAGTTTCTACGATGT
40 AGACTATGTCACTCGGATTGGTAGAGAAGTTATGGAGAGGGAGGAATTTTACCAAGAGATTAGTCGCCATAAGAG
ATTAGTGTGATTTTGGCCCTCAATTGTTACCAGCATTGTTAGAGCATTCTTCTTTTATAATGCCAACTATTTTG
AGGCTTATACAGAGAAGATTATTGACAAAGGTATTAAGCTTTATGAGCGTAATGTTTCCATTATTTAAAGGTTT
TGCCTTATATCAAAAAGGACAGTGTAAAGAAGGCTGTAAGCAGATGCAAGAGGCCATGCATATTTTGTGTTGTT
AGGTCTTCCAGAGCAAGTAGCCTATTATCAGGAACACTACGAAAAATTTGTCAAAGTTAA

45 VGRIRAGVKMEHLGKVFREFRTSGNYSLEAAGESCSTSQLSRFELGESDLAVSRFFEILDNIHVTIENFMDKARNFHN
HEHVSMMQAIIPLYSNDIAGFQKLQREQLKSKSSTPLFYFELNWILLQGLICQDASYDMKQDDLKGKVADYLFKTEE
WTMYELILFNLYSFYDVYVTRJGREVMEREFFYQEISRHKRLVLILALNCYQHCLHSSFYNNANYFEAYTEKIIDKGI
KLYERNVFHYLKGFAlyQKGQCKEQMMEAMHIFDVLGLPEQVAYYQEHYEFVKZ

TABLE 3**ID1 1068bp**

5 ATGTCTAACATTCAAAACATGTCCCTGGAGGACATCATGGGAGAGCGCTTGGTTCGCTACTCCAAGTACATTATTC
AAGACCGGGCTTTGCCAGATATTCGTGATGGGTGGAAGCCGGTTCAGCGCCGTATTCTTTATTCTATGAATAAGGA
TAGCAATACTTTTGACAAGAGCTACCGTAAGTCGGCCAAAGTCAGTCGGGAACATCATGGGGAATTTCCACCCACA
CGGGGATTCTTCTATCTATGATGCCATGGTTCGTATGTACAGAACTGGAAAAATCGTGAGATTCTAGTTGAAATG
10 CACGGTAATAACGGTTCATGGACGGAGATCCTCCTGCGGCTATGCGTTATACTGAGGCACGTTTGTCTGAAATTG
CAGGTACCTTCTCAGGATATCGAGAAAAAGACAGTTCCTTTTGCATGGAACCTTGACGATACGGAGAAAGAAC
CAACGGTCTTGCCAGCAGCCTTTCCAAACCTCTTGGTCAATGGTTCGACTGGGATTTCCGGCTGGTTATGCCACAGA
CATTCTCCCCATAATTAGCTGAGGTCATAGATGTGTCAGTTTACATGATTGACCACCCAACCTGCAAAGATTGAT
AACTCATGGAATCTTGCCTGGACGAGCTTCCTACAGGGGCTATTATTCAGGGTCGTGATGAAATCAAGAAA
15 GCTTATGAGACTGGGAAAGGGCGCGTGGTTGTTCTGTTCCAAGACTGAAATTGAAAAAGCTAAAAGGTGGTAAGGAA
CAAAATCGTTATTATTGAGATTCTTATGAAATCAATAAGGCCAATCTAGTCAAGAAAAATCGATGATGTTCTGTTTA
ATAACAAGGTAGCTGGGATTGCTGAGGTTCTGATGAGTCTGACCGTGATGGTCTTCGTATCGCTATCGAATCTAA
GAAAGACGCTAATACTGAGCTTGTCTCAACTCTTATTTAAGTACACCGACCTACAAATCAACTACAATTTAAT
ATGGTGGCGATTGACAATTTACACCTCGTCAGGTTGGATTGTTCCAATCCTGTCTAGCTATATCGCTCACCCTCG
20 AGAAGTGA

MSNIQNMLEDIMGERFGRYSKYTIQDRALPDIRDGLKPVQRRILYSMNKDSNTFDKSYRKSASVGNIMGNFHPHGD
SIYDAMVRMSQNWKNREILVEMHGNNGSMGDPPAAMRYTEARLSEIAGYLLQDIEKKTVPFAWNFDTEKEPTVLP
AAFPNNLLVNGSTGISAGYATDIPPHNLAVIDAAVYIMDHPTAKIDKLMEFLPGPDFPTGAIHQGRDEIKKAYETGKGRV
25 VYRSKTEIEKLKGGKEQIVIIIEIPYEINKANLVKKIDDVRVNNKVAGIAEVRDESDRDLRIAIELKKDANTELVNLFLK
YTDLQINYNFMVAIDNFTPRQVGLFQSCLAISLTVKEZ

ID12 684bp

30 ATGCCGACATTAGAAATAGCACAAAAAACTGGAGTTCATTAAGAAGGCAGAGAATATTACAATGCCTTGTGT
ACAAATATACAGTTGAGCGGAGATAAACTAAAAAGTAAATTTCCGTTACTCTGTAAACCTGGGGAAGGAAAAACA
ACTACTTCCATAAATATAGCATGGTTCGTTTCCGCGTGCAGGCTATAAACTCTTTTGATCGATGGCGATACTCGAA
ATTCAGTTATGTTAGGAGTTTTTAAATCTCGTGAATAAATTACAGGGCTAACAGAAATTTTATCTGGGACAGCTGA
35 TTTATCTCACGGTTTATGTGATACAAATATTGAAAAATTTTGTAGTTCAATCGGGATCTGTATCACCACCACT
ACAGCCTTGTTACAAAGTAAAAATTTTAAATGATGATTGAAACATTGCGTAAATATTTTGATTATATCATTATTG
ATACACCGCCTATTGGAATTGTTATTGATGCGGCAATTATCACTCAAAGTGTGATGCGTCCATCTTGGTAAACAGC
AACAGGTGAGGCGAATAAACGTGATATCCAAAAAGCGAAACAACAATTAACAAAAACAGGGAACTGTTCTAG
GAGTTGTTTTAAATAAATTGGATATCTCGGTTAATAAGTATGGAGTTTACGGTTCCTATGGAATTTATGGTAAAA
ATAA

40 MPTLEIAQKKLEFIKKAEEYNNALCTNIQLSGDKLVISVTSVNPGEKTTTSINIAWSFARAGYKTLIDGDRNSVML
GVFKSREKITGLTEFLSGTADLSHGLCDTNIENLFVQSGSVSPNPTALLQSKNFNDMIETLRKYFYDIIDTPPIGIVIDAA
IUTQKCDASILVTATGEANKRDIQAKQQLKQTGKLFGLVVLNKLDSVKNYGVYGSYGNYGKKZ

ID13 1182bp

45 ATGGAGGCAAATATGAAACATCTAAAAACATTTTCAAAAAATGGTTTCAATTATTAGTCGTTATCGTCATTAGCT
TTTTAGTGGAGCCTTGGGTAGTTTTTCAATAACTCAACTAACTCAAAAAAGTAGTGTAACAACCTCTAACAACAA
TAGTACTATTACAAAACCTGCCTATAAGAACGAAAAATCAACAACACAGGCTGTAAACAAAGTAAAGATGCTGT
50 TGTTTCTGTTATTACTTATTCGGCAACACAGACAAAATAGCGTATTGGCAATGATGATACTGACACAGATTCTCAG
CGAATCTCTAGTGAAGGATCTGGAGTTATTTATAAAAAAGATGATAAAGAAGCTTACATCGTCACCAACAATCAC
GTTATTAATGGCGCCAGCAAGTAGATATTCGATTGTGATGAGGACTAAAGTACCTGGAGAAATTTGTCGGAGCT
GACACTTTCTGATATTGCTGTCTGCAAAATCTCTTCAGAAAAAGTGACAACAGTAGCTGAGTTTGGTGATTCTA
55 GTAAGTTAACTGTAGGAGAACTGCTATTGCCATCGGTAGCCGTTAGGTTCTGAATATGCAAAATACTGTCACTCA
AGGTATCGTATCCAGTCTCAATAGAAATGTATCCTTAAATCGGAAGATGGACAAGCTATTTCTACAAAAGCCAT
CCAACTGATACTGCTATTAACCCAGGTAACCTCGCGGCCACTGATCAATATTCAAGGGCAGGTTATCGGAAT
TACCTCAAGTAAAAATTGCTACAAATGGAGGAACATCTGTAGAAGGTCTTGGTTTCGCAATTCCTGCAAAATGATGCT
ATCAATATTATTGAACAGTTAGAAAAAACGAAAAAGTGACGCGTCCAGCTTTGGGAATCCAGATGGTTAATTTA
TCTAATGTGAGTACAAAGCAGCATCAGAAAGACTCAATATTCAGTAATGTTACATCTGGTGTAATTGTTCTGTTCCG
60 TACAAAGTAATATGCTGCCAATGGTCACCTTGAATAATACGATGTAATTACAAAAGTAGATGACAAAGAGATTG
CTTCATCAACAGACTTACAAAGTGCTCTTTACAACCACTTATCGGAGACACCACTTAAGATAACCTACTATCGTAA
CGGGAAGAAGAACTACCTCTATCAAACTTAAACAAGAGTTCAGGTGATTAGAATCTTAA

65 MEANMKHLKTFYKKWFQLLVVIVISFFSGALGSFSITQLTKSSVNNNSNNNSTTTQTA YKNENSTTQAVNKVKDAVVSV
ITYSANRQNSVFGNDDTDTDSQRISSESGVYKKNDEAYIVTNNHVINGASKVDIRLSDGTQVPGEIVGADTFSDIAV
VKISSEKVTVAEFGDSSKLTGGETAIAIGSPLGSEYANTVTQGIVSSLNRNVSLKSEDGQAIJTKAIQTDTAIPNGSGGP

LINIQGQVIGITSSKIATNGGTSVEGLGFAIPANDAINIEQLEKNGKVTRPALGIQMVNLSNVSTSDIRRLNIPSNVTSQVIV
RSVQSNMPANGHLEKYDVITKVDDKEIASSTDLSQALYNHSIGDTIKITYRNGKEETTSIKLNKSSGDLESZ

ID15 939bp

5 ATGGCAGAAATTTATCTAGCAGGTGGTTGTTTTGGGGCCTAGAGGAATATTTTTACGCATTTCTGGAGTGCTAG
AAACCAAGTGTGGCTACGCTAATGGTCAAGTCGAAACGACCAATTACCAGTTGCTCAAGGAAACAGACCATGCAG
10 AAACGGTCCAAGTGATTACGATGAGAAGGAAGTGTCACTCAGAGAGATTTACTTTATTATTTCCGAGTTATCGA
TCCTCTATCTATCAATCAACAAGGGAATGACCGTGGTCGCAATATCGAACTGGGATTTATTATCAGGATGAAGC
AGATTTGCCAGCTATCTACACAGTGGTGCAGGAGCAGGAACGCATGCTGGTTCGAAAGATTGCAGTAGAAGTGGA
15 GCAATTACGCCACTACATTCTGGCTGAAGACTACCACCAAGACTATCTCAGGAAGAATCCTTCAGGTTACTGTCA
ATCGATGTGACCGATGCTGATAAGCCATTGATTGATGCAGCAAACTATGAAAAGCCTAGTCAAGAGGTTGTTGAAG
GCCAGTCTATCTGAAGAGTCTTATCGTGTACACAAGAAGCTGTACAGAGGCTCCATTACCAATGCCTATGACC
AAACCTTTGAAGAGGGGATTTATGTAGATATTACGACAGGTGAGCCACTCTTTTTGCCAAGGATAAGTTTGCTTC
15 AGGTTGTGGTTGGCCAAGTTTTAGCCGTCGGATTTCCAAAGAGTTGATTCAATTATACAAGGATCTGAGCCATGGA
ATGGAGCGAATTGAAGTTCGTTCTCGTTCAGGCAGTGTCTCACTTGGGTCTGTTTTCACAGATGGACCGCGGGAGT
TAGGCGGCCTCCGTTACTGTATCAATCTGCTCTTTACGCTTTGTGGCCAAGGATGAGATGGAAAAAGCAGGATA
TGGCTATCTATTGCCTTACTTAAACAAATAA

20 MAEIYLAGGCFWGLEEYFSRISGVLETSVGYANGQVETNNYQLKETDHAETVQVIYDEKEVSLREILLYFRVIDPLSI
NQGNDRGRQYRTGIYYQDEADLPAIYTVVQEQERMLGRKIAVEVEQLRHYLAEDYHQDYLRKNPSGYCHIDVTD
DKPLIDAANYEKPSQEVKASLSEESYRVTEAATEAPFTNAYDQTFEEGIYVDITTEPLFFAKDKFASGCCWPSFSRPI
SKELIHYKDLSHGMERIEVRSRSGAHLGHVFTDGPRELGLRYCINSASLRFVAKDEMEKAGYGYLLPYLNKZ

ID17 870bp

25 ATGAAGATTATTGTACCTGCAACCAGTGCCAATATCGGGCCAGGTTTTGACTCGGTGGTGTAGCTGTAACCAAGT
ATCTTCAAAATGAGGTCTGCGAAGAACGAGATGAGTGGCTGATTGAACACCAGATTGGCAAATGGATTCCACATG
30 ACGAGCGTAATCTTGTCTCAAAATCGCTTTGCAAAATGTACCAAGCTTGCAACCAAGACGCTTGAAATGACCA
GTGATGTCCCTTTGGCGCGCGGTTTGGGTTCTTCAGCTCGGTTATCGTTGCTGGGATTGAAGTACGCCAACCACT
GGGTCAACTCACTTATCAGACCATGAAAAATGCGATTAGCGACCAAGATTGAAGGGCATCCTGACAATGTGGC
TCCAGCCATTATGGTAATCTCGTTATTGCAAGTTCTGTTGAAGGGCAAGTCTCTGCTATCGTAGCAGACTTCCA
35 GAGTGTGATTTTCTAGCTTACATTCAAACTATGAATTACGTACTCGCGACAGCCGTAGTGTCTTGCTAAAAAAT
TGTCTTATAAGGAAGCTGTTGCTGCAAGTTCTATCGCCAAATGTAGCGGTTGCTGCTTGTGGCAGGAGACATGGT
GACCGCTGGGCAAGCAATCGAGGGAGACCTCTCCATGAGCGCTATCGTCAGGACTTGGTAAGAGAATTTGCGAT
GATTAAGCAAGTGACCAAGAAATGGGGCCTATGCAACCTACCTTTCTGGTGTGGGCGGACAGTTATGGTTCT
GGCTTCTCATGACAAGATGCCAACAAATTAAGGCAGAAATGGAAAAGCAACCTTTCAAAGGAAAACCTGCATGACTT
GAGAGTTGATACCAAGGTGTCCGTGTAGAAGCAAAATAA

40 MKIIVPATSANIGPGFDSVGVAVTKYLQIEVCEERDEWLIEHQIGKWIPHDERNLLKIALQIVPDLPRLKMTSDVPLA
RGLGSSSSVIVAGIELANQLGQLNLSDEKLQLATKIEGHPDNVAPAIYGNLVIASSVEGQVAIVADFPCEDFLAYIPNY
ELRTRDSRSVLPKKLSYKEAVAASSIANVAVAALLAGDMVTAGQATEGDLFHERYRQDLVREFAMIKQVTKENGAYAT
YLSGAGPTVMVLASHDKMPTIKAELEKQPFKGLHDLRVDTQGVREAKZ

ID20 564bp

45 ATGAAATATCAGGATTACATCTGGGATTTAGGTGGAACCTTACTGGATAATTATGAAACTTCAACAGCTGCATTTG
TTGAAACATTGGCACTGTATGGTATCACACAAGACCATGACAGTGTCTATCAAGCTTTAAAGGTTTCTACTCCTTT
50 TGGCATTGAGACATTGCTCCCAATTTAGAGAATTTTTAGAAAAGTACAAGGAAAATGAAGCCAGAGAGCTTGA
ACACCCGATTTTATTGAAGGAGTTTCTGACCTATTGGAAGACATTTCAAATCAAGGTGGCCGTCATTTTTTGGTC
TCTCATCGAAATGATCAGGTTTTGGAAATTTAGAAAAAACCCTATAGCAGCTTATTTTACAGAAAGTGGTGACTT
CTAGCTCAGGCTTTAAGAGAAAGCCAAATCCCGAATCCATGCTTTATTTAAGAGAAAAGTATCAGATTAGCTCTG
55 GTCTTGTGATGATCGCCGATTGATATCGAAGCAGGTCAAGCTGCAGGACTTGATACCCACTTGTTTACCAG
TATCGTGAATTTAAGACAAGTATTAGACATATAA

60 MKYHDYIWDLGGLLDNYETSTAFAFVETLALYGITQDHSVYQALKVSTPFAIETFAPNLENLEKYKENEARELEHPI
LFEGVSDLLEDISNQGRHFLVSHRNDQVLEILEKTSIAAYFTEVVTSSSGFKRKPENPESMLYLRKYQISSGLVIGDRPID
IEAGQAAGLDTHLFTSIVNLRQVLDIZ

ID21 1875bp

65 ATGACAGAAGAAATCAAAAATCTGCAGGCACAGGATTATGATGCCAGTCAAATTCAGTTTATGAGGGCTTAGAG
GCTGTTTCGATGCGTCCAGGGATGTACATTGGATCAACCTCAAAAAGGAAGGCTTCCACCATCTAGTCTGGGAAATG
TTGATAACTCAATTGACGAGGCCTTGGCAGGATTTGCCAGCCATATTCAAGTTTTATTGAGCCAGATGATTCGAT
TACTGTTGTGGATGATGGGCGTGGTATCCAGTCGATATTCAGGAAAAAACAGGCCGCTCTGCTGTTGAGACCGT

CTTTACAGTCCTTCACGCTGGAGGAAAGTTCCGGCGGTGGTGATACAAGGTTTCAGGTGGTCTTCACGGGGTGGG
GTCGTCAGTAGTTAATGCCCTTTCCACTCAATTAGACGTTTCATGTTCAAAAAATGGTAAGATTCATTACCAAGAA
TACCGTCGTGGTCATGTTGTCGAGATCTTGAAATAGTTGGAGATACGGATAAACAGGAACAACTGTTCACTTC
ACACCGGACCCAAAAATCTTCACTGAAACAACAATCTTTGATTTTGATAAAATTAATAAACGGATTCAAGAGTTG
5 GCCTTTCTAAATCGCGGTCTTCAAATTTCAATTACAGATAAGCGCCAAGGTTTGGAAACAAACCAAGCATTATCATT
ATGAAGGTGGGATTGCTAGTTACGTTGAATATATCAACGAGAAACAAGGATGTAATCTTTGATACACCAATCTATA
CAGACGGTGAGATGGATGATATCACAGTTGAGGTAGCCATGCAGTACACAACCTGTTACCATGAAAAATGTCATGA
GTTTCGCCAATAATATTCATACCCATGAAGGTGGAAACACATGAACAAGGTTTCCGTACAGCCTTGACAGCTGTTAT
CAACGATTATGCTCGTAAAAATAAGTTACTGAAAGACAATGAAGATAATTTAACAGGGGAAGATGTTCCGGAAGG
10 CTTAACTGCAGTTATCTCAGTTAAACACCCAAATCCACAGTTTGAAGGACAAACCAAGACCAAAATGGGAAATAG
CGAAGTGGTCAAGATTACCAATCGCCTCTTCAGTGAAGCTTTCTCCGATTTCCTCATGGAATAATCCACAGATTGCC
AAACGTATCGTAGAAAAAGGAATTTTGCTGCCAAGGCTCGTGTGGCTGCCAAGCGTGCAGTGAAGTCAACAGT
AAAAATCTGTTTGGAAATTTCCAACTTCCAGGGGAACTAGCAGACTGTTCTTCTAATAACCCGTGCTGAAACAG
15 AACTCTTCACTCGTGAAGGAGACTCAGCTGGTGGATCAGCCAAATCTGTCGTAACCGTGAGTTTCAGGCTATCCT
TCCAATTCGCGGTAAAGATTTTGAACGTTGAAAAAGCAAGTATGGATAAGATTCTAGCCAACGAAGAAATTCGAG
TCTTTTACAGCCATGGGAACAGGATTTGGCGCAGAAATTTGATGTTTCGAAAGCCGTTACCAAAAACTCGTTTGG
ATGACCGATGCCGATGTCGATGGAGCCACATTCTGACCTTCTTTAACCTTGATTATCGTTATATGAAACCAA
TCCTAGAAGCTGGTTATGTTTATATTGCCCAACCACCAATCTATGGTGTCAAGGTTGGAAGCGAGATTAAGAATA
TATCCAGCGGGTGCAGATCAAGAAATCAAATCCAAGAAGCTTTAGCCCGTTATAGTGAAGGTCGTACCAACC
20 GACTATTACGCGTTATAAGGGGCTAGGTGAAATGGACGATCATCAGCTGTGGGAAACAAACCATGGATCCCGAACA
TCGCTTGATGGCTAGAGTTTCTGTAGATGATGTGCAGAAGCAGATAAAATCTTTGATATGTTGA

MTEEIKNLQAQDYDASQIQVLEGLEAVRMRPGMYIGSTSKEGLHHLVWEIVDNSIDEALAGFASHIQVFIEPDDSTIVVD
DGRGIPVDIQEKTGRPAVETVFTVLHAGGKFGGGYKVSGLHGVSSVVALSTQLDVBHVKNGKIHVQEYRRGHV
25 VADLEIVGDTDKTGTTHFTPDPKIFTEITTFDFDLNKRQELAFNRLQISTDKRQGLEQTKHYHYEGGIAHYVEYI
NENKDVIFDTPITDGEDDITVEVAMQYTTGYHENVMSFANNIHTHEGGTHEQGFRTALTRVINDYARKNKLKDN
EDNLTGEDVREGLTAVISVKHPNPQFEGQTKLGNSEVVKITNRLFSEAFSDFLMENPOIAKRIVEKILAAKARVAAC
RAREVTRKSGLEISNLPGKLADCSNNPAETELFIVEGDSAGGSAGSRNREFQAILPIRGKILNVEKASMDKILANEEI
30 RSLFTAMGTGFGAEFDVSKARYQKLVMTDADVDGAHIRTLLLTLYRYMKPILEAGYVYIAQPPYGVKVGSEIKEYI
QPGADQEIKLQELALARYSEGRTKPTIQRYKGLGEMDDHQLWETTMDEHRLMARVSVDDVQKQKSLICZ

ID54 1446bp

ATGAGTAGACGTTTTAAAAATCACGTTACAGAAAGTGAAAGCGAAGTGTTAATATAGTTTTGCTGACTATTTATT
TATTGTTAGTTTGTGTTTTTATTGTTCTTAATCTTAAGTACAATATCCTTGCTTTTAGATATCTTAATCTAGTGGTAA
35 CTGCGTTAGTCTACTAGTTGCCTTGGTAGGGTACTCTTGATTATCTAAAAAAGCTGAAAAGTTTACTATTTTT
CTGTTGGTGTCTCTATCCTTGTGAGCTCTGTGCTGCTCTTGCAGTACAGCAGTTTGTGGAGTACCAATCGTTT
AAATGCGACTTCTAATTACTCAGAATATTCAATCAGTGTGCTGTTTTAGCAGATAGTGAGATCGAAAATGTTACG
CAACTGACGAGTGTGACAGCAGCAGTGGGACTAATAATGAAAATATTAGAAAATTAAGCTGATATCAAGTCA
40 AGTCAGAATACCGATTGACGGTCAACAGAGTTCGTCTTACTTGGCAGCTTACAAGAGTTTGATTGCAGGGGAG
ACTAAGGCCATTGTCCTAAATAGTGTCTTTGAAACATCATCGAGTCAAGTATCCAGACTACGCATCGAAGATA
AAAAAGATTTATACTAAGGGATTCACTAAAAAAGTAGAAGCTCCTAAGACGCTAAGAGTCAGTCTTTCAATATC
TATGTTAGTGGAAATTGACACCTATGGTCTTATAGTTCCGGTGTGCGGATCAGATGTCAACATCCTGATGACTGTCA
45 ATCGAGATACCAAGAAAATCCTCTTGACCACAACGCCACGTGATGCTATGTACCAATCGCAGATGGTGGAAATA
ATCAAAAAAGATAAATTGACTCATGCGGGCATTTATGGAGTTGATTGCTCCATTACACCTTAGAAAAATCTCTATGG
AGTGGATATCAATTACTATGTGCGATTGAACCTTCACTTCGTTTTGAAATTGATTGATTTGTTGGTGGAATTGATG
TTTATAATGATCAAGAATTTACTGCCATACGAATGGAAGTATTACCTGCAGGCAATGTTTCATCTTGATTCAGA
ACAGGCTCTCGGTTTTGTTGAGCGCTACTCCCTAGCAGATGGCGATCGTGACCGCGGGCGCCATCAACAAAA
50 GGTGATTGTGGCTATCCTTCAAAAAATTAACTCAACCGAAGTGTGAAAAATTATAGTACGATCATTAATAGCTTT
CAAGATTCTATCCAAACAAATATGCCACTTGAGACCATGATAAATTTGGTCAATGCTCAGTTAGAAAGTGGAGGG
AATTATAAAGTAAATTCTCAAGATTAAAGGGACAGGTCGGATGGATCTTCCTTCTTATGCAATGCCAGACAGTA
ACCTCTATGTGATGGAATAGATGATAGTAGTTTAGCTGTAGTTAAAGCAGCTATACAGGATGTGATGGAGGGTA
GATGA

MSRRFKKRSQKVKRSVNIVLLTIYLLLVCFLLFIKYNILAFRYLNLVVTALVLLVALVGLLLIYKKAKEFTIFLLVFS
ILVSSVSLFAVQQFVGLTNRLNATSNYSEYSIVAVLADSEIENVTLQTSVTAAPTGTNNENIQKLLADIKSSQNTDLTVNQ
SSSYLAAKYSLIAGETKAIVLNSVFENIESEYDPYASKIKKIYTKGFTKKVEAPKTSKSQSFNIYVSGIDTYGPISSVSRSDV
NILMTVNRDTKILLTTTPRDAYVPIADGGNNQKDKLTHAGIYGVDSSTHTLENLYGVVDINYVRLNFTSFLKLIDLLGG
IDVYNDQEFTHANGKYYPAGNVHLDEQALGFVRERYSLADGDRDRGRHQKQVIVAILQKLTSTEVLNKYSTIINSLO
60 DSIQTNMPLETMINLVNAQLESGGNYKVNQDLKGTGRMDLPSYAMPDSNLYVMEIDSSLA VVKAAIQDVMEGRZ

ID55 732bp

ATGATAGACATCCATTGCGATATCGTTTTTGATGTAGATGACGGTCCCAAGTCAAGAGAGGAAAGCAAGGCTCTC
TTGGCAGAATCCTACAGACAGGGGGTGGCAACCATTTGTTTCTACCTCTCACCGTCGCAAGGGCATGTTTGAAACTC
CGGAAGAGAAGATAGCAGAAAACTTTCTTCAGGTTCCGGAAATAGCTAAGGAAGTGGCGAGTGACTTGGTCATTG
5 CTTACGGGGCTGAAATTTATTACACACCAGATGTTCTGGATAAGCTGGAAAAAAGCGGATTCCGACCCTCAATG
ATAGTCGTTATGCCTTGATAGAGTTTAGTATGAACACTCCTTATCGCGATATTCATAGCGCCTTGAGCAAGATCTT
GATGTTGGGAATTACTCCAGTCATTGCCACATTGAGCGCTATGATGCTCTTGAAAAAATGAAAAACGGCTTGA
GAACTGATCGATATGGGCTGTTACACGCAAGTAAATAGTTACATGTCCTCAAAACCCAACTTTTGGCGAAGCTT
10 ATAAATTCATGAAAAAAGAGCTCAGTATTTTTAGAGCAGGATTGGTTCATGTCATTGCAAGTGATATGCACAA
TCTAGACGGTAGACCTCCTCATATGGCAGAAGCATATGACCTTGTACCCAAAAATACGGAGAAGCGAAGGCTCA
GGAACTTTTTATAGACAATCCTCGAAAAATTGTAATGGATCAACTAATTTAG

MIDIHSHIVFDVDDGPKSREESKALLAESYRQGVRTIVSTSHRRKGMFETPEEKIAENFLQVRELAKEVASDLVIA YGAEI
YYTPDVLDKLEKKRIPTLNDSTRYALIEFSMNTPYRDIHSLSKILMLGITPVIAHIERYDALENNEKRVRELIDMGCTYQV
15 NSSHLVKPKLFGERYKFMKKRAQYFLEQDLVHVIASDMHNLDRPPHMAEA YDLVTQKYGEAKAQELFIDNPRKIVM
DQLIZ

ID58 3990bp

TTGATTTATATAATCGCTATCAATATAACAATGCAATCAGGAGGTTTTGCAATGAAACATGAAAAACAACAGCGT
TTTTCTATTGTAATACGCTGTAGGAGCAGCTTCTGTTCTAATTTGGATTGCTTCCAAGCAGAGCTGTTGCGA
20 CCGATGGAGTTACTCTACTACTACAGAAAAACCAACCGACCATCCATACGGTTTTCTGATTCCCTCAATCATCCGA
AAATCGGACTGAGGAAACACCTAAAGCAGTGCTTCAACCAGAAGCTCCAAAACTGTAGAAACAGAACTCCAG
CTACTGATAAGGTAGCTAGTCTTCCAAAAACAGAAAGAAACCAAGAGGAAGTTAGTTCAACTCCTAGTGATA
AAGCAGAAGTGGTAACCTCAACTTCTGCTGAAAAAGAACTGCTAATAAAAAAGGCAGAAAGCTAGCCCTAAA
25 AAGGAAGAAGCGAAAGAGGTTGATTCTAAAGAGTCAATACAGACAAGACTGACAAGGATAAACCAAGCTAAAA
AGATGAAGCGAAAGCAGAGGCTGACAAACCGGCAACAGAGGCAGGAAAGGAACGTGCTGCAACTGTAATGAAA
AACTAGCGAAAAAGAAAATTGTTTCTATTGATGCTGGACGTAATATTTCTCACCAGAACAGCTCAAGGAAATCA
TCGATAAAGCGAAACATTATGGCTACACTGATTTACACCTATTAGTCGGAAATGATGGACTCCGTTTCATGTTGGA
CGATATGAGCATCAGCTAACGGCAAGACCTATGCCAGTGACGATGTCAAACGCGCCATTGAAAAAGGTACAAA
30 TGATTATTACAACGATCCAAACGGCAATCACTTAACAGAAAGTCAAATGACAGATCTGATTAATATGCCAAAGA
TAAAGGTATCGGTCTCATTCCGACAGTAAATAGTCTGGACACATGGATGCGATTCTCAATGCCATGAAAGAATT
GGGAATCCAAAAACCTAATCTTAGCTATTTTGGGAAGAAATCAGCCCGTACTGTGCTGATCTTGACAACGAACAAGC
TGTCGCTTTTACAAAAGCCCTTATCGACAAGTATGCTGCTTATTTCCGCGAAAAAGACTGAAATCTTCAACATCGGA
35 CTTGATGAATGCCAATGATGCGACAGATGCTAAAGGTTGGAGTGTGCTTCAAGCTGATAAATACTATCCAAAC
GAAGGCTACCTGTAAAAGGCTATGAAAAATTTATGCTACGCCAATGACCTCGCTCGTATTGTAATAATCGCAC
GGTCTCAAACCAATGGCTTTAACGACGGTATCTACTACAATAGCGACACAAGCTTTGGTAGTTTGGACAAAGAC
ATCATCGTTTCTATGTGGAAGTGGTGGTGGGAGGCTACGATGTGCTTCTTAACTACTAGCTGAAAAAGGTC
ACCAATCCTTAATACCAATGATGCTTGGTACTACGTTCTTGGACGAAACGCTGATGGCCAAGGCTGGTACATCT
40 CGATCAGGGGCTCAATGGTATTAACCAACACCAATCACTTCTGTACCAAAAAACAGAGGAGCTGATATCCCAAT
CATCGGTGGTATGGTAGCTGCTTGGGCTGACACTCCATCTGCACGTTATTACCATCACGCCTCTTCAAACTCATG
CGTCATTTTGCAATGCCAAGCTGAATACTTCGACGCTGATTATGAATCTGCAGAGCAAGCACTTAACGAGGTA
CCAAAAGACCTGAACCGTTATACTGCAGAAAGCGTACCGCCGTAAGAAAGCTGAAAAAGCTTTCGCTCTCTC
45 GATAGCAACCTTAGCCGTGCCCAACAGATACGATTGATCAAGCCATTGCTAACTTCAAGAACTGTCAACAAC
TTGACCCTCACGCCTGAAGCTCAAAAAGAGAAGGCTAAACGCTGAGGTTGAAAAACTTGCCAAAAACAAGGT
AATCTCAATCGATGCTGGACGCAATACTTTACTCTGAACACGCTCAAACGCATCGTAGACAAGGCCAGTGAGCT
CGGATATTTGATGTCCATCTCTTCTAGGAAATGACGGACTTCGCTTTCTACTCGATGATATGACCATTACTGCC
AACGGAACCACTATGCTAGTGATGACGTTAAAAAAGCTATTATCGAAGGAACTAAAGCTTACTACGACGATCCA
50 AACGGTACTGCACTAACACAGGCAGAAAGTAACAGAGCTAATGAATACGCTAAATCTAAGGACATCGGTCTCATC
CCAGCTATTAACAGTCCAGGTACATGGATGCTATGCTGGTTGCCATGGAAAAATTAGGTATTAAAAATCTCTCA
GCCCACCTTTGATAAAGTTTCAAAAACACTATGGACTTGAAAAACGAAGAGCGATGAACCTTTGTAAGGCCCTC
ATCGGTAAATACATGGACTTCTTTGACGTTAAAAACAAAGATTTTCACTTTGGTACTGACGAATACGCCAACGAT
GCGACTAGTGCCCAAGGCTGGTACTACCTCAAGTGGTATCAACTCTATGGCAATTTGCCGAATATGCCAACACC
55 CTCGCAGCTATGGCCAAAGAAAGAGGGCTTCAACCAATGGCCTTCAACGATGGCTTCTACTATGAAGACAAGGAC
GATGTTTCAGTTTGACAAAGATGTCTTGAATTTCTACTGGTCTAAAGGCTGGTGGGGATATAACCTCGCATCACTC
AATACCTAGCAAGCAAGGCTATAAATCTTGAATACCAACGGTGACTGGTACTACATTCTTGGTCAAAAAACAG
AAGATGGTGGTGGTTTCTCAAGAAAGCTATTGAGAATACTGGAAAAACACCACTTCAATCAACTAGCTTCTACCA
AATATCTCTGAAGTAGATCTTCCAACAGTCGGAAGTATGCTTCAATCTGGGCAGATAGACCAAGCGCTGAATACA
AGTAAGAGGAAATCTTTGAACCTCATGACTGCCTTTGCAGACCACAACAAAGACTACTTTCGTGCTAATTAATG
60 CTCTCCGGGAAGAAATAGCTAAAATCTTCAAACTTAGAAGGATATAGTAAAGAAAGCTTGTAGGCCCTTGACG
CAGCTAAAAACAGCTCTAAATTACAACCTCAACCGTAATAAACAAGCTGAGCTTGACACGCTTGTAGCCAACCTAA
AAGCCGCTCTTCAAGGCTCAAAACAGCTGTAACCTATTTCAGGAAGCTAGATGAAAAATGAAGTGCTGCCAATG
TTGAAACCAAGACCAAGAACTCATCAAGAACTGAAGAAATTCATTGAAAGTTATCAAGAAAGAAAAATCTTAACC
TCCAGCCGGTCAGGAAAAATATTATCACAGCAGGAGTCAAAGGTGAACGAACCTATTACATCTCTGTACTCACTG
65 AAAATGGAAAAACAACAGAAACAGTCTTGATAGCCAGGTAACCAAGAAAGTTATAAACCAAGTGGTTGAAGTT
GGCGCTCCTGTAACCAAGGGTGATGAAGTGGTCTTGACCAACTACTGAGGTAAAAACCTAGACTGGATATC

5 CAAGAAGAAGAAATTCCATTTACCACAGTGACTTGTGAAAATCCACTCTTACTCAAAGGAAAAACACAAGTCATT
ACTAAGGGCGTCAATGGACATCGTAGCAACTTCTACTCTGTGAGCACTTCTGCCGATGGTAAGGAAGTAAAAACA
CTTGTAATAAGTGTGCTAGCACAGGAAGCGTTACTCAAATAGTGAAGTCGGAAGTATGGTAACACATGTAGGC
GATGAAAAACGACAGCCGCTATTGCTGAAGAAAAACCAAACTAGAAATCCCAAGCCAACCAAGCTCCATCAAC
TGCTCCTGCTGAGGAAAGCAAAGTTCTTCTCAAGATCCAGCTCCTGTGTTAACAGAGAAAAAACTTCTGAAAC
AGGAACTCAGGATTCTGCAGGACTAGTAGTCGAGGACTCATGTCCACACTAGCAGCCTATGGACTCACTAAAAAG
AAAAGAAGACTAA

10 MIYIAINITMQSGGFAMKHEKQQRFSIRKYAVGAASVLIGFAFQAQTVAADGVPTTTTENOPTIHTVSDSPQSENRTTE
TPKAVLQPEAPKTVEETPATDKVASLPKTEEPQEEVSTPSDKAEVVTPTSAREKETANKKAEESPKEEAEKEVDSKE
SNTDKTDKDKPAKKDEAKAEADKPATEAGKERAATVNEKLAKKKTVIDAGRKYFSPEQLKEIIDKAKHYGYTDLHLL
VGNDGLRFMLDDMSITANGKTYASDDVKRAJEKGTNDYNDPNGNHLTESQMTDLINYAKDKGIGLIPTVNSPGHMD
AILNAMKELGIQNPNSFYFGKKSARTVDLDNEQAVAFKALIDKYAAYFAKKTETIFNIGLDEYANDATDAKGWSVLQA
15 DKYYPNEGYPVKGYEKFLAYANDLARIVKSHGLKPMFNDGIYNSDTSFGSFDKDIIVSMWTGGWGGYDVASSKLLA
EKGHQILNTNDAWYYVLGRNADGGQWYNLDQGLNGIKNTPTSPVKTEGADIPIGGMVAAWADTPSARYSPSRLFKL
MRHFANANAAYFAADYESAEQALNEVPKDLNRYTAESVTAVKAEKAIRSLDSNLSRAQQDITDQALAKLQETVNNLT
LTPEAQKEEEEAKREVEKLAKNKVISIDAGRKYFTLNQLKRIVDKASELGYSDVHLLGNDGLRFLDDMTTANGKTYA
20 SDDVKKAIIEGTKAYYDDPNGTALTQAEVTELIEYAKSKDIGLIPAINSPGHMDAMLVAMEKLGIKNPQAHFDKVSMTT
MDLKNEEAMNFKALIGKYMDFFAKGTKIFNFGTDEYANDATSAQGWYLLKWWYLYGKFAEYANTLAAMAKERGL
QPMFNDGFYFEDKDDVQFDKDVLSYWSKGWGYNLASPYLASKGYKFLNTNGDWYLLGQKPEDGGGFLKKAI
ENTGKTPFNQLASTKYPEVDLPTVGSMLSIWADRPSEYKEEEIFELMTAFADHNKDYFRANYNAREELAKIPTNLEG
YSKESLEALDAAKTALNYNLNRNKQAEIDTLVANLKAALQGLKPAVTHSGSLDENEVAANVETRELPETITREIEPFEVI
25 KKENPNLPAGQENITAGVKGERTHYISVLTENGKTTETVLDQVTKVINQVVEVGAPVTHKGDESGLAPTEVVKPRL
DIQEEIEPFTTVCENPLLLKGKTQVITKGVNGHRSNFYSVTSADGKEVKTLVNSVVAQEAQVQVEVGTMTVTHVGDE
NGQAAIAEEKPKLEIPSPAPSTAPAEESKVLQDPAPVVTTEKKLPETGTHDSAGLVVAGLMSTLAAAYGLTKRKEDZ

ID122 825bp

30 ATGAACAAAAACAAGACAGACACTAATCGGACTGCTAGTGTTATTGCTTTTGTCTACAGGGAGCTATTATATC
AAGCAGATGCCGTCGGCACCTAATAGTCCCAAAACCAATCTTAGTCAGAAAAACAAGCGTCTGAAGCTCCTAGT
CAAGCATTGGCAGAGAGTGTCTTAACAGACGCAAGTCAAGAGTCAAATAAAGGGGAGTCTGGAGTGGAAATGGCTC
AGGTGCTTTTATCGTCAATGGTAATAAAACAAATCTAGATGCCAAGGTTTCAAGTAAGCCCTACGCTGACAAATA
35 AACAAAGACAGTGGGCAAGGAACTGTTCAACCGTAGCTAATGCCCTCTGTCTAAGGCCACTCGTCAGTACAA
GAATCGTAAAGAACTGGGAATGGTCAACTCTTGGACTCCTCCAGGTGGCATCAGGTCAAGAATCTAAAGGG
CTCTTATACCCATGCAGTCGATAGAGGTCAATTGTTAGGCTATGCCTTAATCGGTGGTTGGATGTTTGTATGCTC
CAACAAGCAATCTAAAAACATTGCTGTTCCAGACAGCCTGGGCAATCAGGCACAAGCCGAGTATTGACTGGTC
AAAACTACTATGAAAGCAAGGTGCGTAAAGCCTTGGACCAAAACAAGCGTGTCCGTTACCGTGTAAACCCCTTACT
ACGCTTCAACGAGGATTAGTTCCTCAGCTTACAGATTGAAGCAAGTCTTCGGATGGAGAATTGGAATTCA
40 ATGTTCTAGTTCCTCAATGTTCAAAAGGGGACTTCAACTGGATTACCGAACTGGAGAAGTAAGTCAACTCAGTAA

45 MNKKTROTLLIGLLVLLLLSTGSYYIKQMPSPNPKTNLSQKKQASEAPSQALAESVLTDAVKSQIKGSLEWNGSGAFTV
NGNKTNLDAKVSSKPYADNKTCTVGKETVPTVANALLSKATROKYKRNKETGNGSTSWTPPGWHQVKNLKGSYTHAV
DRGHLLGYALIGGLDGFDASTSNPKNIAVQTAWANQAQAEYSTGQNYYESKVRKALDQNKRVRYRVTLYYASNEDL
VPSASQIEAKSSDGELEFNVLVPNVQKGLQLDYRTGEVTVTQZ

ID123 225bp

50 GTGCTAAGATTCAGCGGATTGAGGCAAGTGATGAAGATGAATAAGAAATCAAGCTACGTAGTCAAGCGTTTACTT
TTAGTCATCATAGTACTGATTTTAGGTACTCTGGCTCTAGGAATCGGTTTAAATGGTAGGTTATGGAATCTTGGGCA
AGGGTCAAGATCCATGGGCTATCCTGTCTCCAGCAAAATGGCAGGAATTGATTCATAAATTACAGGAAATTAG

VLRFSGLRQVMKMNKSSYVVKRLLLVIIVLILGTLALGIGLMVGYGILGKGQDPWAILSPAKWQELIHKFTGNZ

55

CLAIMS:

1. A *Streptococcus pneumoniae* protein or polypeptide having a sequence selected from those shown in table 1.
5
2. A *Streptococcus pneumoniae* protein or polypeptide having a sequence selected from those shown in table 2.
3. A protein or polypeptide as claimed in claim 1 or claim 2 provided in substantially pure form.
10
4. A protein or polypeptide which is substantially identical to one defined in any one of claims 1 to 3.
5. A homologue or derivative of a protein or polypeptide as defined in any one of claims 1 to 4.
15
6. An antigenic and/or immunogenic fragment of a protein or polypeptide as defined
20 in Tables 1-3.
7. A nucleic acid molecule comprising or consisting of a sequence which is:
25
 - (i) any of the DNA sequences set out in Table 1 or their RNA equivalents;
 - (ii) a sequence which is complementary to any of the sequences of (i);

- (iii) a sequence which codes for the same protein or polypeptide, as those sequences of (i) or (ii);
- (iv) a sequence which is substantially identical with any of those of (i), (ii) and (iii);
- (v) a sequence which codes for a homologue, derivative or fragment of a protein as defined in Table 1.
8. A nucleic acid molecule comprising or consisting of a sequence which is:
- (i) any of the DNA sequences set out in Table 2 or their RNA equivalents;
- (ii) a sequence which is complementary to any of the sequences of (i);
- (iii) a sequence which codes for the same protein or polypeptide, as those sequences of (i) or (ii);
- (iv) a sequence which is substantially identical with any of those of (i), (ii) and (iii);
- (v) a sequence which codes for a homologue, derivative or fragment of a protein as defined in Table 2.
9. The use of a protein or polypeptide having a sequence selected from those shown in Tables 1-3, or homologues, derivatives and/or fragments thereof, as an immunogen and/or antigen.

10. An immunogenic and/or antigenic composition comprising one or more proteins or polypeptides selected from those whose sequences are shown in Tables 1-3, or homologues or derivatives thereof, and/or fragments of any of these.
- 5 11. An immunogenic and/or antigenic composition as claimed in claim 10 which is a vaccine or is for use in a diagnostic assay.
12. A vaccine as claimed in claim 11 which comprises one or more additional components selected from excipients, diluents, adjuvants or the like.
- 10 13. A vaccine composition comprising one or more nucleic acid sequences as defined in Tables 1-3.
14. A method for the detection/diagnosis of *S.pneumoniae* which comprises the step of bringing into contact a sample to be tested with at least one protein or polypeptide as defined in Tables 1-3, or homologue, derivative or fragment thereof.
- 15 15. An antibody capable of binding to a protein or polypeptide as defined in Tables 1-3, or for a homologue, derivative or fragment thereof.
- 20 16. An antibody as defined in claim 15 which is a monoclonal antibody.
17. A method for the detection/diagnosis of *S.pneumoniae* which comprises the step of bringing into contact a sample to be tested and at least one antibody as defined in claim 15 or claim 16.
- 25 18. A method for the detection/diagnosis of *S.pneumoniae* which comprises the step of bringing into contact a sample to be tested with at least one nucleic acid

sequence as defined in claim 7 or claim 8.

19. A method of determining whether a protein or polypeptide as defined in
Tables 1-3 represents a potential anti-microbial target which comprises inactivating
5 said protein or polypeptide and determining whether *S.pneumoniae* is still viable.

20. The use of an agent capable of antagonising, inhibiting or otherwise
interfering with the function or expression of a protein or polypeptide as defined in
Tables 1-3 in the manufacture of a medicament for use in the treatment or
10 prophylaxis of *S.pneumoniae* infection

1 / 2

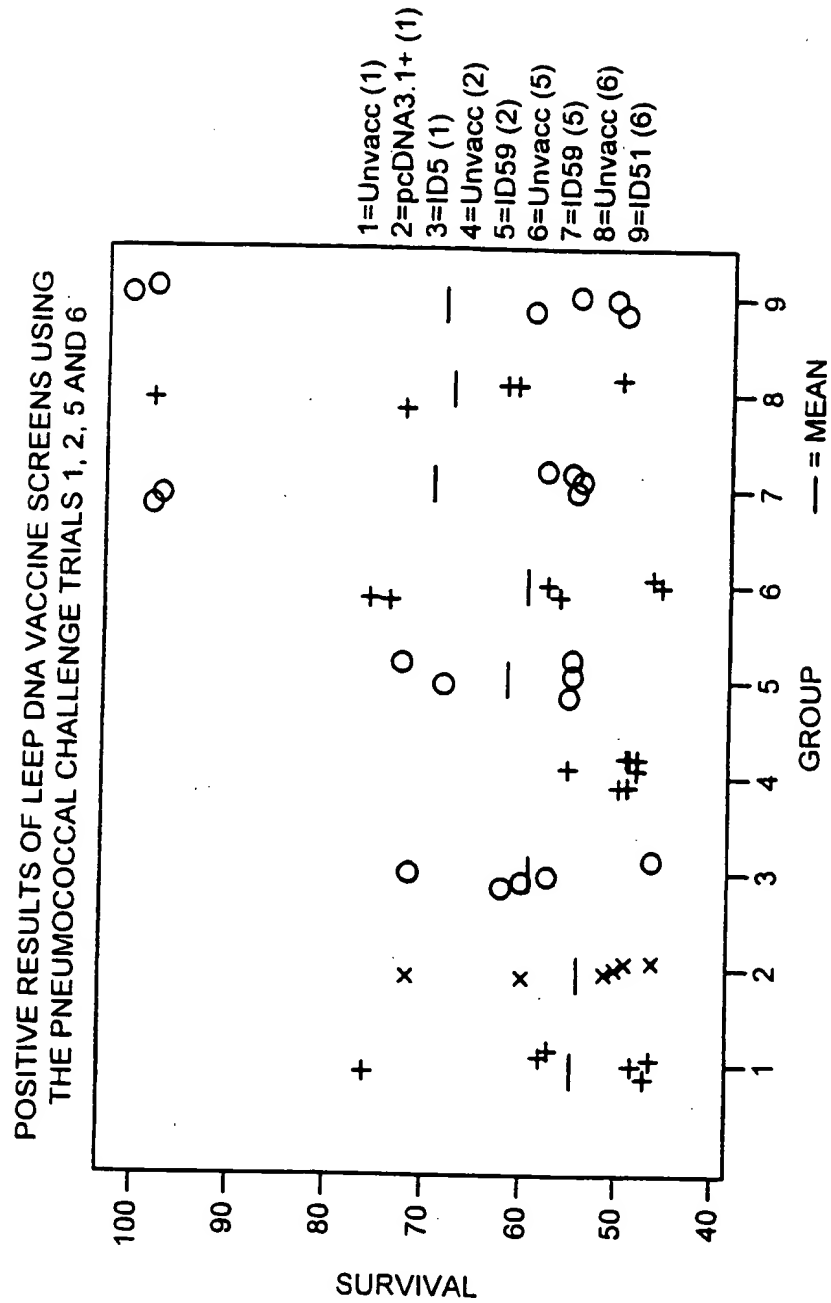


FIG. 1

2 / 2

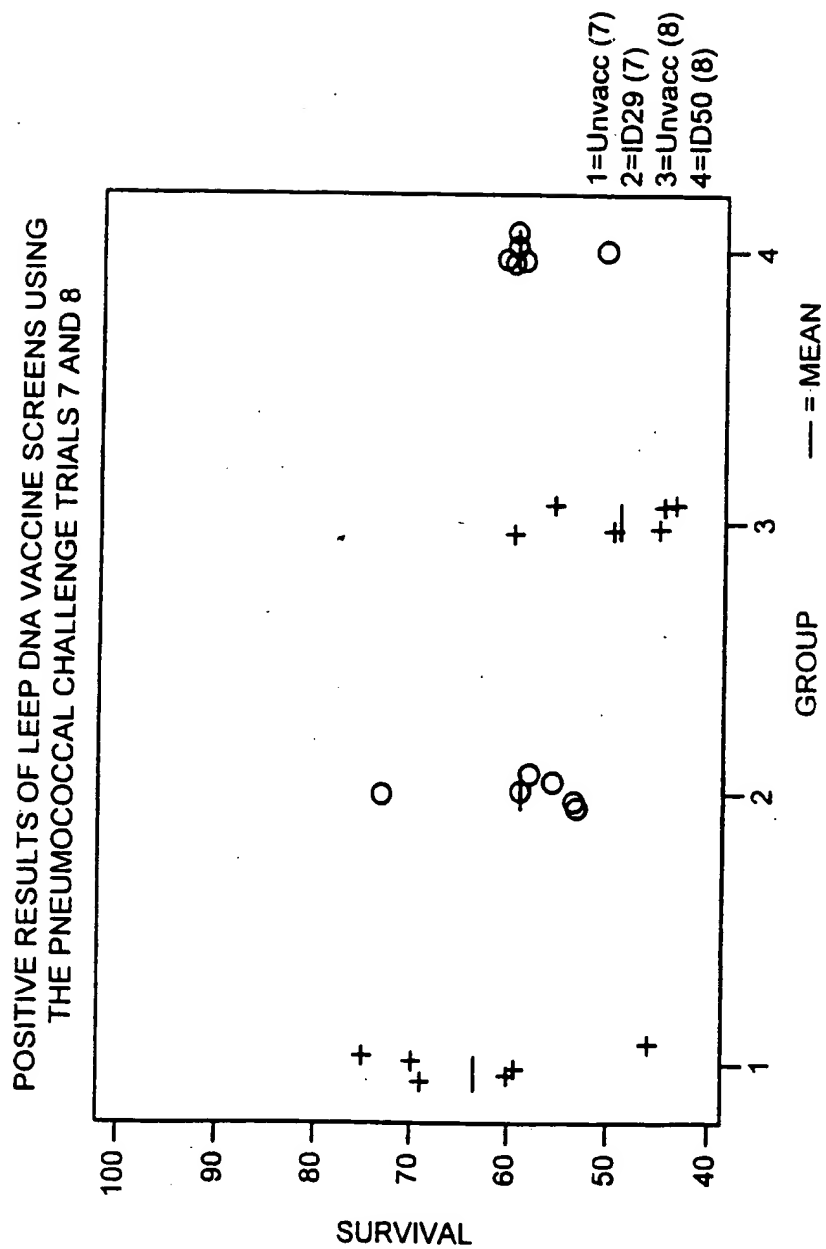


FIG. 2